Essays on the Spatial Distribution of Economic Activities

Yi Jie Gwee

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

Abstract

Essays on the Spatial Distribution of Economic Activities

Yi Jie Gwee

This dissertation consists of three chapters that examine the spatial distribution of economic activities. The first chapter, co-authored with Jacob Field and Di Song Tan, examines how disasters as well as individuals' expectations of what others will do affect the development of cities. The development of cities often involves the rejuvenation or replacement of existing structures. However, history, in the form of the sunk cost of existing durable structures, often serves as an impediment to urban development. In theory, by reducing the opportunity cost of waiting to rebuild to zero, disasters can eliminate these frictions and bring about higher quality structures. In addition, the simultaneous rebuilding after a disaster would allow property owners to experience stronger cross-building spillovers which would encourage further upgrades of nearby buildings. Nevertheless, these are not sufficient to guarantee higher quality buildings. This is because individuals' investment decisions also depend on their expectations of what others will do. Therefore, in this chapter, we examine both of these issues using the 1666 Great Fire of London as a natural experiment. First, using a difference-in-differences (DiD) strategy, we show evidence that the Fire was able to free parishes within London from the constraints of their existing durable structures and move them to a new equilibrium involving higher quality structures. Second, using DiD and an IV strategy, we find that legal rulings arising from the Fire Court – a court specially set up by the English Parliament to hear rebuilding disputes – were able to anchor expectations and in so doing, helped to facilitate the development of London. Providing causal evidence that legal rulings can be a main driver in the formation of expectations is the main contribution of our paper.

The second chapter examines how the quirks of history shape present-day economic outcomes. Building on Bazzi et al. (2020), I study how a particular episode of history – time at the frontier – helps to explain the present-day manufacturing production patterns across American counties. First, I show empirical evidence that there are fewer establishments and lower employment in counties that

spent a longer time on the frontier. The same results hold for industries that are more "contractible" (i.e., easier to specify in contracts and hence less susceptible to holdup). Second, using a DiD strategy, I show that firms in high "contractibility" industries sort into producing at counties that spent a longer time on the frontier. I hypothesize that due to "rugged individualism", individuals in counties that spent a longer time on the frontier are less likely to trust other people. Therefore, anything that is not "contractible" becomes harder and more costly to enforce. Consequently, only the more "contractible" industries locate in counties that spent a longer time on the frontier.

The third chapter, co-authored with Di Song Tan, examines how land use regulations and NIMBY ("not in my back yard") behavior affect housing prices in the UK. In the UK, developers have to apply to the local planning authority to seek development permission. Applicants who have their plans rejected can appeal to the Secretary of State, via the Planning Inspectorate. The Planning Inspectorate then assigns an inspector to decide whether to overturn the local authority's decision. We propose a theoretical model which shows that in locations with high levels of NIMBY-ism, developers are better off getting their plans rejected by the local authority and gambling on drawing an inspector who is less sympathetic towards locals' NIMBY behavior. Our empirical strategy exploits the fact that inspectors are quasi-randomly assigned to the appeals. This allows us to use inspector leniency as an instrument for whether an appeal is successful. We find that overturning the local authority's decision does not lead to a large fall in housing prices. For some projects, the impact may in fact be positive because they also add to local amenities such as retail shops. This suggests a prevalence of NIMBY-ism, as locals pressure authorities to reject even relatively benign projects.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This dissertation would not have been possible if not for the guidance and support of many. I am greatly indebted to my advisor Donald Davis. Despite his busy schedule, Don took the time to meet me regularly. Our meetings would often last for hours even when they took place on the weekend. He has been my fiercest critic yet my kindest mentor, constantly challenging me to do better but encouraging me every step of the way. His advice has been instrumental in all three chapters of this dissertation, and more importantly, in the way that I think about economics and life. I appreciate his clarity in approaching questions and his sense of humor which never fails to make a seemingly helpless situation hopeful. Don kept believing in my potential even in the moments where I ceased to do so. I will always be thankful for the tremendous support and patience that he has shown me – I could not have asked for a better advisor.

I am also grateful to my other committee members, Réka Juhász, David Weinstein, Michael Best and Suresh Naidu, for their advice, comments, as well as encouragement. I am thankful for the numerous opportunities that Réka has given me. From having a Skype call with me when I was first accepted into Columbia, to hiring me as a research assistant at the end of my first year, to taking me on board as her teaching assistant in my third year, and to supporting my applications for academic awards and research grants, Réka has been extremely supportive throughout my time at Columbia. David has always asked thought provoking questions and helped me to sharpen what exactly are my research questions. Michael has been super helpful in constantly being able to suggest ways to solve identification issues and for always being able to suggest a helpful reference. I thank Suresh for opening my eyes to the literature on the intersection of political economy and economic history.

Special mention must go to Amy Devine for often being the unsung hero of the department. Amy has been unbelievably helpful with all the administrative paper work. I would also like to thank Jacob Field for so generously sharing data on the hearth tax and Fire Court judges that we use in the first chapter of this dissertation.

My fellow PhD classmates Howard Zhang, Ken Hong Teoh and Sakshi Gupta made the PhD journey a lot more enjoyable. Having spent more than three years out of school, the intensity of the first year was difficult. Despite always being at least ten steps behind them, Howard, Ken and Sakshi were extremely patient and took the time to explain the difficult concepts to me – often at the expense of their own revision time. I am also thankful for their friendship outside of the classroom and will always treasure our meet-ups for coffee or a meal. I have also benefited from conversations with Naman Garg, Yu Kyung Koh and Pablo Warnes.

For unwavering friendship and for reminding me that there is life outside of research, I thank Russell Low, Ruth Teng, Di Song Tan, Hui Ren Tan and Royston Toh. To Di Song and Hui Ren, I have benefitted tremendously from our conversations and will always marvel at how we can come up with research ideas even when we are trying our best not to talk about research. I hope that this would not mark the end but the start of many more papers that we will write together.

I was first introduced to economic research as an undergraduate at the LSE where I had the wonderful fortune of being able to interact extensively with Steve Pischke and Judith Shapiro. Judith recognized my interest in research and played a key role in encouraging me to apply to graduate school. Steve has continued to have an important influence in the way that I approach applied work and this is something that I will always be thankful for. I have also been greatly inspired by Steve's commitment to teaching.

Prior to embarking on the PhD at Columbia, I had spent a few years at the Ministry of Trade and Industry, Singapore and the Ministry of Social and Family Development, Singapore. The work and the people that I met were instrumental to my development as an economist. I would like to specially thank Andy Feng, Suan Wee Foo, Chun Seng Yip and Yik Wei Yong for mentoring me and for moving mountains to get me the necessary approvals to embark on the PhD even though I

had yet to complete my existing bond. For financial support, I would like to thank the Ministry of Trade and Industry, Singapore for awarding me the PhD scholarship.

Last but definitely not least, I am deeply grateful for my family. My parents have sacrificed much so that my sisters and I can pursue our dreams. From young, they taught us how to be inquisitive, engaged us in deep discussions and encouraged us to pursue our dreams. I thank my sisters for always being able to make me laugh. My one big regret is spending so many years half a world away from home, and in so doing, missed the precious growing up years of my youngest sister. I thank my in-laws for showing genuine interest in my work and for their prayers. Special thanks to my cousin Lily and Mateo, and their kids Ilya and Eva, for inviting us over for dinners, board game nights and basketball afternoons. They have allowed me to experience the warmth of family although I was miles away from home. I am beyond grateful for Lily, Mateo and Louis, who despite the own difficulties that they were facing, helped us to pack and ship everything back to Singapore when our time in New York was suddenly cut short due to Covid.

Above all, my deepest gratitude is to my wife Jiahui who has been incredibly supportive even though it meant giving up countless opportunities of her own. In the midst of the soul crushing moments as a researcher, I know that I can always count on her encouragement and support. Even when I lacked a good identification strategy, a model that cannot be closed or simply a boring research question, Jiahui has accepted me for all my limitations, loving me for who I am and supporting me in every way possible.

To paraphrase Isaac Newton, I am where I am not because I have stood on the shoulders of giants. Instead, it is because they have, by grace, lifted me onto their shoulders, so that I can tiptoe, and glimpse just a little further.

Dedication

All cities are transient, except for the one to come.

To all who have ever had to say goodbye to someplace, something or someone.

Chapter 1: Great Expectations: Urban Development in 17th Century London (with Jacob Field and Di Song Tan)

1.1 Introduction

The development of cities often involves the rejuvenation or replacement of existing structures. However, history, in the form of the sunk cost of existing durable structures, often serves as an impediment to urban development. In every period, property owners face a trade-off between receiving rent from the existing building or incurring a cost to tear down the building and rebuilding it. As a result, they often wait long periods of time for their building to depreciate before embarking on upgrading. Furthermore, without some gain to being the first to upgrade their property, property owners may rationally wait for others to upgrade first. In theory, by reducing the opportunity cost of waiting to rebuild to zero, disasters (such as a Fire) can eliminate these frictions and bring about higher quality structures. In addition, the simultaneous rebuilding after a disaster would allow property owners to experience stronger cross-building spillovers. As described by Hornbeck and Keniston (2017), this "virtuous circle" of cross-plot externalities results in building upgrades encouraging further upgrades of nearby buildings.

Nevertheless, the opportunity cost of waiting to rebuild falling to zero coupled with the prospects of stronger cross-building spillovers, are not sufficient to guarantee higher quality buildings. This is because individuals' investment decisions also depend on their expectations of what others will do. For example, if a city (or more generally, an area) is growing, then individuals will expect other individuals to build higher quality buildings. By contrast, if the expectations are that the area is in decline, then individuals may not even rebuild or may invest at a lower quality since they expect other individuals to do the same.

Therefore, in this paper, we examine both of these issues using the 1666 Great Fire of London as a natural experiment. Our research questions are as follows. First, we examine whether the Fire was able to free parishes[1] within London from the constraints of their existing durable structures and move them to a new equilibrium involving higher quality structures. In line with the historical context, we define the quality of structures based on the number of hearths in the property. While the first research question that we examine is similar to the papers on the 1872 Boston fire by Hornbeck and Keniston (2017) and the 1906 San Francisco fire by Siodla (2015), our second research question departs from these papers. In particular, we study what anchors individuals' expectations of what others will do and how this can consequently facilitate the development of cities. We find evidence that legal rulings arising from the Fire Court – a court specially set up by the English Parliament to hear rebuilding disputes – were able to anchor expectations and in so doing, helped to facilitate the development of London. Providing causal evidence that legal rulings can be a main driver in the formation of expectations is the main contribution of our paper.

For the first part of the paper, to examine whether the removal of development frictions through the Fire resulted in higher quality structures being rebuilt, we employ a difference-in-differences (DiD) strategy. The DiD strategy exploits both the cross-sectional and time-series variations arising from the Fire. The time-series variation comes from the timing before and after the Fire which was exogenous. The cross-sectional variation arises because different parishes in London were affected differently by the Fire. For example, some parishes were burned whereas some parishes did not experience any damage from the Fire at all. A null effect from our regression would suggest that there were no frictions to upgrading before the Fire – the quality of properties was optimal. By contrast, a positive effect suggests the presence of upgrading frictions which the Fire effectively removed.

[1] Parishes were administrative units within a city that played a role in both civil and ecclesiastical matters.

Using our DiD strategy, we find that a few years after the Fire, burned parishes experienced a highly statistically significant increase in the number of hearths per property compared to unburned parishes. In addition, the effect varied with the level of damage. Parishes which were more badly damaged saw a highly statistically significant increase in the number of hearths per property compared to parishes which were less damaged. Finally, the effect was biggest for parishes whose neighboring parishes were all burned compared to parishes whose neighboring parishes were not all burned.

The result from the first part of the paper suggests that individuals had positive expectations that others will be rebuilding at a high quality. Nevertheless, it does not tell us what is driving these expectations. Therefore, in the second part of the paper, we examine the role of legal rulings in driving expectations. In 17th century England, tenants were legally obliged to rebuild in the event of any disasters which damaged the property, even if it was not their fault. However, the Fire took place amidst a plague and war – an unprecedented joint occurrence of events. To expedite the rebuilding of London, the English Parliament established the Fire Court.

The second part of the paper begins with a model that shows that legal rulings affect expectations because they affect the bargaining between landlords and tenants who do not go to Court. This is because their outside options are based on the Fire Court's initial rulings. For our empirical strategy, we turn once again to a DiD strategy. Just as before, the time-series variation comes from the timing before and after the Fire. However, the cross-sectional variation now arises because different parishes experienced different Fire Court rulings. For example, some parishes saw a disproportionate number of initial cases where the Fire Court voided the existing contracts between the landlord and tenant and consequently assigned the rebuilding to the landlord. This is what we refer to as pragmatic rulings. Voiding the contract means that both the landlord and tenant surrender their contracts. This allows both parties to negotiate a new contract with each other or other parties.

Our regression results show that parishes with a greater share of pragmatic rulings had more hearths per property compared to parishes where there was a lower share of cases with pragmatic rulings. In addition, because only a very small proportion of properties in each parish went to the Fire Court, our results suggest that the rulings of these few cases had an outsized effect on the quality of other buildings in the parish.[2] Why would this be the case? We argue that this is because the small share of cases was enough to anchor expectations.

While we have included a number of time varying parish-level controls in our regression, a threat to identification in the DiD strategy is that we might not have controlled for all possible confounders. As a result, the change in the number of hearths may be related to changes in parish level characteristics that are not due to the Fire Court rulings – a violation of the parallel trend assumption. Therefore, we augment our DiD strategy with an instrumental variable (IV) strategy.

Our IV strategy exploits the fact that at the parish level, Fire Court judging panels that have different political alignments (i.e., whether they were predominantly Royalists or Parliamentarians) were assigned to the cases. The 1666 Great Fire took place in the midst of the Second Dutch War (1665-1667) and the Great Plague which began in 1665. King Charles II was relying on loans from London and its wealthiest citizens to finance the war. The destruction of the customs house, wharves and more than 13,000 buildings caused a significant drop in royal revenue. The King had a vested interest for London to be rebuilt quickly. Therefore, judging panels that consisted predominantly of Royalists (i.e., more aligned with the King) were more likely to decree pragmatic rulings so as to facilitate the rebuilding of London. As a result, we can use the composition of the judging panels as an instrument for the share of cases in the parish that had pragmatic rulings. This gives us exogenous variations in legal rulings for each parish.

[2] Based on the initial cases, the average proportion of properties in each parish that went to the Fire Court was 6%.

We find that the results from our IV analysis re-affirm our DiD results – legal rulings can indeed anchor expectations and help to facilitate the rebuilding process. To the best of our knowledge, while there are theoretical papers such as Cooter (1998), Basu (2000), McAdams (2000, 2005), Myerson (2004) and Hadfield and Weingast (2012) that examine how legal institutions can affect expectations and hence the behavior of individuals, there are relatively fewer empirical papers that provide causal evidence of this.

In examining how expectations affect the behavior of economic agents, our paper is related to Krugman (1991) and Rauch (1993). In addition, our paper is related to how cities recover from major shocks and whether they move to a new equilibrium. Beginning with Davis and Weinstein (2002), there has been an extensive literature that examines whether long-run city size is robust to temporary shocks. These shocks include wars and bombing (Davis and Weinstein (2002) and Miguel and Roland (2011)), natural or man-made disasters (Siodla (2015) and Hornbeck and Keniston (2017)), political events (Redding et al. (2011) and Michaels and Rauch (2018)), technology (Bleakley and Lin (2012)) and even diseases (Jedwab et al. (2019)). Our paper provides evidence of how the Great Fire of London freed London from the constraints of history and enabled it to move to a new equilibrium with more hearths per property.

By addressing how legal rulings contribute to the development of cities, our paper is related to the literature on the economic consequences of legal origins. This literature shows how legal origins affect particular legal rules and these in turn affect economic outcomes such as growth, financial development, property rights and contract enforcement. Examples of these studies include Acemoglu et al. (2001), Djankov et al. (2003), La Porta et al. (1997), La Porta et al. (1998) and Dell (2010). In using judging panels that consisted predominantly of Royalist as our instrument in our IV analysis, our paper is also related to North and Weingast (1989), Acemoglu et al. (2005), Jha (2015) and Angelucci et al. (2020). These papers examine the tensions between Parliamentarians and Royalists during various times in English history (e.g., the English Civil War (1642-1651) and the Glorious Revolution (1688)) and show how these affected the development of institutions that facilitated growth in England.

Finally, our paper is related to the historical literature on the impact of the Great Fire of London. Field (2008) notes that the 1666 Great Fire of London is such an iconic moment in the history of London that the contemporary media frequently used the phrase "The Second Great Fire" to describe the London Blitz during World War II. While the 1666 Fire has been extensively studied by historians (e.g., Reddaway (1940), Porter (1996) and Field (2018)) and even legal scholars (e.g., Tidmarsh (2016)), our paper contributes to this largely qualitative literature by providing a quantitative analysis on the impact of the Great Fire of London.

The rest of the paper proceeds as follows. Section 1.2 presents the historical background of the 1666 Great Fire of London. Section 1.3 discusses the novel data sources that we use for our analysis. Section 1.4 examines the effect that the Fire had on the quality of properties that were rebuilt. Section 1.5 presents our main contribution which is that legal rulings anchored individuals' expectations of what others will do and this consequently facilitated the development of parishes within London. We conclude in Section 1.6.

1.2 Historical Background: The 1666 Great Fire of London

This section draws extensively from Reddaway (1940), Porter (1996), Field (2008), Tidmarsh (2016) and Field (2018). The Great Fire of London began on September 2, 1666, in a bakery on Pudding Lane in the City of London. The City of London covers an area of 2.8 km^2 or 1.1 $miles^2$ within London and was home to about one sixth of London's inhabitants. The structure of the city made it easy for the Fire to spread. Streets, lanes and alleys were narrow and buildings were made from timber. In addition, the upper floors of houses often cantilevered over the pathways below. This meant that the top floors on one side of the street nearly touched those on the other side, making it easy for the Fire to spread. The Fire lasted for three and a half days and destroyed approximately 13,200 buildings in the City of London. An estimated 70,000 out of 80,000 inhabitants living in the City of London lost their homes.

Tidmarsh (2016) notes that despite the urgency to rebuild London, there were significant challenges. At the time of the Fire, the institution of fire insurance had not yet developed. Instead, the common practice was that leases had a covenant that obligated the tenant, regardless of whether the tenant was at fault, to repair or rebuild the premises in the event of disasters or wars. This created substantial challenges for both the tenants and landlords. For the tenants, there was the issue of fairness in whether they should bear the full cost of rebuilding. Many tenants could not afford to rebuild. Moreover, tenants who had a short time left on their lease had little incentive to rebuild. As for the landlords, there were long delays and huge cost in bringing disputes to the common-law courts. Even if the case was brought before the common-law courts, the powers of these courts were constrained by the existing tenancy agreements. As a result, the judges could not calibrate or void the existing contracts to achieve the best incentives for the parties to rebuild. Furthermore, due to the existing tenancy agreements, landlords could not prematurely re-enter the leased premises in order to facilitate reconstruction.

In order to expedite the rebuilding of London, the English Parliament established the Fire Court to adjudicate between landlords and tenants as to who would bear the burden of rebuilding (Fire of London Disputes Act 1666). The bill was passed in the House of Lords on January 23, 1667. A few days later, on January 31, 1667, the House of Commons assented to the bill.[3] Tidmarsh (2016) notes that the Fire Court heard a total of 1,585 cases. Some cases involved more than one property so the 1,585 cases understate the extent of the Court's work. As mandated by the Fire Court legislation, each case was heard by a panel of at least three judges. The judges were given the

[3]The year of the enactment of the statute was listed as 1666 even though the bill was passed in the House of Lords on January 23, 1667. This is because based on the calendar that was used during that era, the new year began on Lady Day (March 25).

power to void existing contracts and decide the details of the new contracts (e.g., who rebuilds, new rent and length of the tenancy agreement, etc.). The typical process when a case is brought to the Fire Court is that the judges would first try to mediate and get the tenant and landlord to come to an agreement. In the event that the parties are unable to come to an agreement, the Court will then make a ruling which is legally binding.[4]

In concluding our discussion about the historical background of the Great Fire of 1666, we would like to highlight that there were previously other fires in London that also resulted in substantial damage. For example, Richardson (2001) notes that the Great Fire of 1133 damaged St Paul's, St Bride's, London Bridge and properties as far east as Aldgate. Another example was the Great Fire of 1212 which began at Southwark, destroyed the church and spread to London Bridge. Legal issues surrounding the responsibility of the tenant to rebuild would have also existed back then. Why then was the Fire Court only set up after the 1666 Fire and not earlier? The existing literature is surprisingly silent on this.

One reason could be that while previous Great Fires caused substantial damage, the damage to property from the 1666 Fire was arguably the greatest (see for example Garrioch (2016)). London had grown substantially since the 12th and 13th century. Therefore, even if the entire city was almost destroyed due to the 1133 Fire, by 1666 the size of the city would have been far larger. Nevertheless, due to the lack of data (most of the evidence is qualitative), it remains debatable whether the damage from the 1666 Fire was the greatest. For example, Garrioch (2016) notes that about 3,000 people died

[4]Reddaway (1940), Porter (1996) and Tidmarsh (2016) note that besides setting up the Fire Court, the Parliament of England also put in place other legislation and measures to facilitate the rebuilding of the city. There were new building regulations to limit damage from subsequent fires (buildings must be made of brick, be of a minimum size, not exceed a certain height and must not cantilever over the streets). To determine boundaries and settle disputes among neighbors, a survey system was put in place. Since properties were taken for public purposes (e.g., widened streets), there was a formal channel to value property. To finance the reconstruction of public buildings, a tax on coal was introduced. To ensure that property owners rebuilt within a reasonable time, sanctions were meted out if this was not done. There were provisions requiring owners to share rebuilding costs that benefited multiple properties (e.g., party walls). Regulations on the price and quality of raw materials used for the rebuilding were implemented. Incentives were given to encourage skilled craftsmen to come to London to help with the rebuilding.

in the 1212 Fire, far more than the eight people that was estimated to have died due to the 1666 Fire.[5]

Therefore, we think that the main reason was due to the joint occurrence of war, plague and Fire – a combination of events that was absent in the previous Great Fires. The Great Plague which began in 1665 resulted in the death of almost a quarter of London's population within 18 months. This means that there was now a huge excess supply of vacant properties which vastly increased the bargaining power of tenants. The King could wait for the landlords and tenants to reach a bargained outcome. For example, whether the landlord contributes to the rebuilding or changes the terms of the tenancy contract even though by law the tenant has to rebuild. However, given the ongoing Second Dutch War (1665 to 1667), King Charles II simply could not wait for this to play out. Tidmarsh (2016) argues that the King was relying on loans and taxes from London and its wealthiest citizens to finance the war. The destruction of the customs house, wharves and buildings caused a significant drop in royal revenue from custom and hearth taxes. The Fire Court was therefore a way to expedite reaching a somewhat equitable outcome. It gives the landlord class some portion of what prior precedent would suggest but it also tilts things sufficiently toward tenants to mirror the shift in bargaining power owing to the plague.[6]

1.3 Data

Urban investment. In line with how the value of a property was assessed in 17th century London, we measure quality by the number of hearths that are in each property before and after the Fire. This information is available from the historical manuscripts of the hearth tax assessment records that are held at The National Archives, United Kingdom.

[5]Despite the destruction, the largest estimate of deaths directly due to the Fire was eight. This is a shockingly small number and historians such as Field (2018) have offered a number of explanations. First, the incineration of bodies in the Fire meant that corpses could not be recovered and so the death records are underestimates. Second, the Fire took place over three and a half days. This gave sufficient time for people to evacuate. Third, historians postulate that the relatively tight-knit nature of the neighborhoods meant that there was help and assistance for the vulnerable.

[6]We would like to thank Don Davis for helping us to sharpen this argument.

According to the University of Roehampton, Centre for Hearth Tax Research,[7] the hearth tax was introduced in England and Wales in 1662 to provide a regular source of income for King Charles II who was the newly restored monarch. Parliament had estimated that the King required an annual income of £1.2 million. However, by 1661, there was a shortfall of £300,000 and it was hoped that the hearth tax would make up for this. The hearth tax was essentially a property tax on dwellings graded according to the number of fireplaces in the property. The tax was paid in two equal installments at Michaelmas (September, 29) and Lady Day (March, 25) by the occupier. If the property was vacant, the landlord paid the tax. In order to administer the tax, a list of householders was compiled and this formed the hearth tax assessment records.

Our pre-Fire hearth data come from two sources. First, we use the full records from the 1666 London and Middlesex hearth tax, along with portions of the 1663 and 1664 documents that have been cleaned and digitized by the London Hearth Tax project.[8] Since the hearth tax was collected twice a year in March and September, the 1666 records are based on the March collection which took place before the Fire in September. Second, we supplement this with the 1664/1665 Southwark hearth tax records that come from the assessment for Surrey. The Southwark data were manually transcribed by Field (2008) for his history PhD thesis.[9]

As for the post-Fire hearth data, we rely on the records from the 1675 London and Middlesex hearth tax records as well as the 1673 Surrey (Southwark) hearth tax records. These data were also

[7]https://www.roehampton.ac.uk/research-centres/centre-for-hearth-tax-research/

[8]In June 2007, the London Hearth Tax project was formed to systematically analyze and digitize the hearth tax records. The project united the expertise of the British Academy Hearth Tax Project, the Centre for Hearth Tax Research (University of Roehampton), Birkbeck College (University of London), and the Centre for Metropolitan History (Institute of Historical Research). In 2011, the full records from the 1666 London and Middlesex hearth tax, along with portions of the 1663 and 1664 documents, were published electronically via British History Online at https://www.british-history.ac.uk/london-hearth-tax/london-mddx/1666.

[9]While the data from Southwark is undated, Field (2008) notes that they are most certainly from the period between 1664 and 1665.

manually transcribed by Field (2008).[10]

The unit of geography for our analysis is at the parish level. Due to the differences in the scope and range of the hearth tax assessments, some parishes only appear in the pre-Fire records while others only appear in the post-Fire records. In our regressions, we only use data from the parishes that appear in both the pre- and post-Fire records. Table 1.1 shows the summary statistics of the hearth tax data which we use in our regressions.

Table 1.1: Summary statistics (Hearth Tax data)

	Mean	SD	Min	Max	N
Number of hearths (pre-Fire)	3.83	3.79	0	193	44,724
Number of hearths (post-Fire)	4.33	3.36	0	135	35,006
Number of parishes	.	.	.	.	70

Some might question whether the number of hearths is a reasonable way to measure the quality of the building. We believe that it is reasonable for a few reasons. First, unlike assessed values or market values, the number of hearths is an objective measure and is not based on a valuation. Second, Field (2008) documents that research has shown that there is some correlation between the number of hearths and wealth, as well as occupation. To the extent that the wealthier and those with higher social standing live in higher quality buildings, then we should expect the number of hearths to be a reasonable proxy for the quality of the building.

Details of Fire Court judges. The Fire Court was composed of England's twelve common-law judges. There were three common-law courts (Common Pleas, King's Bench, and the Exchequer) with four justices appointed to each court. In the years after the Fire, some judges retired or passed

[10]Although the London and parts of the Middlesex hearth tax records were presented to Parliament sessions on February 1, 1675, Field (2008) states that a faded note on the manuscript linked it to a collection on 1674. Other parts of the Middlesex records were based on an assessment between 1674 and 1675. The data for Southwark come from an assessment for Surrey that was not dated. Field (2008) notes that it is probably associated with a collection in 1673.

away and hence our sample contains fourteen judges and not twelve.

Table 1.2: Summary statistics (Judges)

	Mean	SD	Min	Max	N
Year of birth	1603.14	6.79	1587	1611	14
Year called to bar	1629.14	6.51	1614	1637	14
Year knighted	1658.71	7.02	1643	1668	14
Pro-restoration of monarchy	.57	.43	0	1	14
Pro-established church	.57	.43	0	1	14
Studied at Oxford University	.36	.5	0	1	14
Served in Grays Inn	.07	.27	0	1	14
Served in Lincolns Inn	.36	.5	0	1	14
Served in Inner Temple	.5	.52	0	1	14
Served in Middle Temple	.07	.27	0	1	14
From Common Pleas	.43	.51	0	1	14
From King's Bench	.29	.47	0	1	14
From Exchequer	.29	.47	0	1	14
Head of common-law Court	.29	.47	0	1	14
Number of judges	.	.	.	.	14

In order to get details about the Fire Court judges, we referred to various sources such as the Oxford Dictionary of National Biography (2004) and Sainty (1993). From these sources, we obtained information on the judges. Many seismic political events took place in 17th century England. For example, the English Civil War (1642-1651), the restoration of the monarchy (1660), as well as the Puritans' (English Protestants) continuous attempts to get the Church of England (established church) to abandon its Roman Catholic practices. Therefore, from these books, we also obtained information on the judges' religious views and their views on the 1660 restoration of the monarchy (i.e., whether they were Royalists or Parliamentarians). We define binary variables for whether the judges were supportive of the restoration of the monarch (Royalists) and whether they were supportive of the established church. We assign the value of 0.5 if the judges had moderate views. In our IV analysis, we use the composition of the judging panels an instrument for the share of initial cases in the parish that had pragmatic rulings.

Table 1.2 shows us the summary statistics of the Fire Court judges. On average, the judges tend to be slightly pro-restoration of the monarchy and pro-established church. Around 36% of the judges attended Oxford University with the rest attending Cambridge University. The majority of the judges trained at the Inner Temple. Finally, 43% of the Fire Court judges were from the court of the Common Pleas and 29% of them were the respective heads of their common-law courts (i.e., Lord Chief Justice or Lord Chief Baron).

Details of Fire Court cases. The transcripts of the cases that were heard by the Fire Court were compiled into nine volumes. These records survive up to today and are housed at the London Metropolitan Archives. To commemorate 300 years since the Fire, in 1966, four volumes (volumes A, B, C and D) were calendared (summarized) and converted to modern English by Philip E. Jones. These were subsequently published as two books – Jones (1966, 1970). The summaries contain extremely detailed information. For example, they give us details on who the landlords and tenants are, the location of the property, the rent and tenure of the tenancy contract before the Fire, the day that the case was heard by the Fire Court, the judges who heard the case, as well as the new rent and tenure that were decreed by the panel of judges. Figure A.1 shows how some of the case characteristics evolved over time (within the first 716 days).

As part of his history PhD thesis, Field (2008) transcribed some of the information associated with the cases in these four volumes into a dataset. We augment this dataset by transcribing additional information that was not captured by Field (2008). Table 1.3 shows us the summary statistics of the Fire Court data based on the cases that we have sufficient information. In 13% of the cases in our sample, the Fire Court voided the existing contracts (i.e., both landlord and tenant surrendered the existing contract) and assigned the cost of rebuilding to the landlord. In 1.3% of the cases, the judges altered the existing contracts (i.e., no surrendering) and assigned the rebuilding to the landlord. In 71% of the cases, the judges altered the existing contracts and assigned the rebuilding to the tenant. In 10% of the cases, the Fire Court voided the existing contracts but decreed the sharing of cost in the rebuilding. Finally, in 5.2% of the cases, the judges altered the existing contracts but decreed the sharing of cost in the rebuilding. The fine paid is the lump-sum payment made on execution of the lease. For each judging panel, we calculate the share of judges that were supportive of the established church and the share of judges that were supportive of the 1660 restoration of the monarchy (Royalists). On average, in each judging panel, 48% of the judges tend to be supportive of the restoration of the monarchy and 46% tend to be supportive of the established church. This suggests that the judging panels were on average quite moderate in their views.

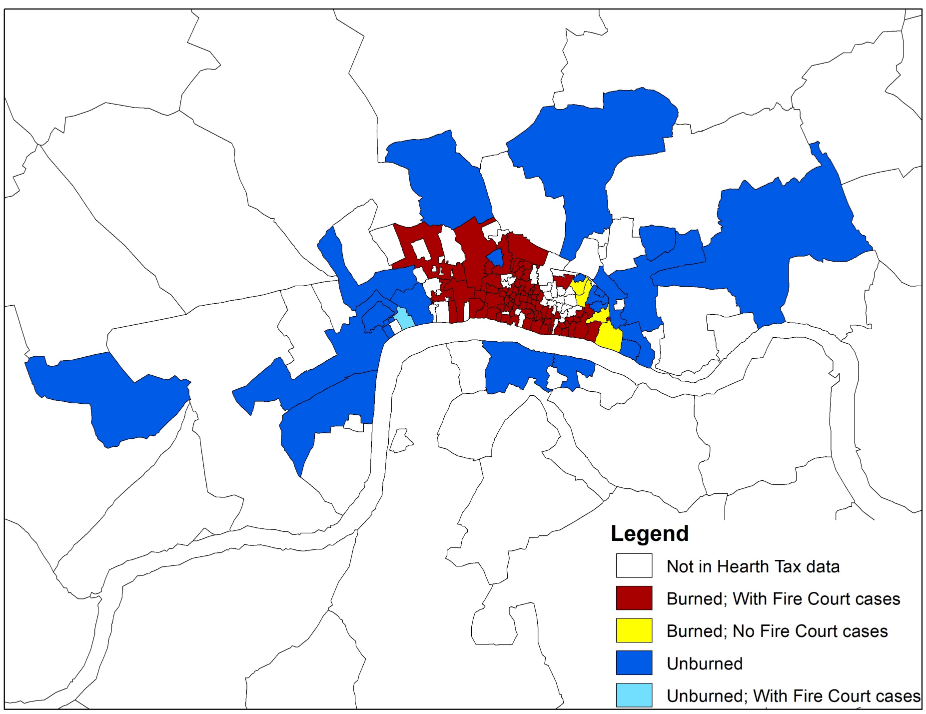
Table 1.3: Summary statistics (Fire Court data)

	Mean	SD	Min	Max	N
Both parties surrender: Owner rebuilds	.13	.33	0	1	696
No surrender: Owner rebuilds	.01	.11	0	1	696
No surrender: Tenant rebuilds	.71	.46	0	1	696
Both parties surrender: Cost sharing	.10	.30	0	1	696
No surrender: Cost sharing	.05	.22	0	1	696
Degree of separation from owner	1.18	.52	1	6	696
Start year of tenancy	1655.74	10.23	1591	1666	679
Years left in tenancy	34.75	387.83	0	9996.92	663
Fine paid	69.08	202.61	0	4000	696
Rent per annum	31.29	36.91	0	474	692
Amount spent on improvements	48.46	233.07	0	3000	696
Average pro-monarchy of panel	.48	.22	0	1	696
Average pro-church of panel	.46	.21	0	1	696
Head of common-law Court	.57	.6	0	3	696
Number parishes	.	.	.	.	67
Number cases	.	.	.	.	696

Finally, since we are interested in examining whether legal rulings of the initial cases in each parish can anchor expectations, we do not actually need to observe the rulings of all the cases that went to the Fire Court. Therefore, the four out of nine volumes which have been calendared would suffice for our purposes as these four volumes cover the earlier cases. We refer to these cases from the first four volumes as the "initial" cases.

Regression sample. Putting all our data sources together, using the shapefiles provided by Satchell et al. (2018), Figure 1.1 shows the parishes that are included in our regressions. In the diagram, we label a parish as "burned" as long as any part of it was damaged by the Fire.

Figure 1.1: Regression sample



1.4 The Effect of the Fire

The 1666 Great Fire of London had both quantity and quality effects on the development of London. On the quantity side, the Fire affected the total number of properties and hearths in each parish. As for quality, the Fire affected the number of hearths per property in each parish. In this paper, we focus on the effect that the Fire had on quality. This is because the plague wiped out about a quarter of London's population. Therefore, we should expect fewer properties to be rebuilt in the immediate aftermath since there are now fewer people to house. However, the effect on quality is not clear. In addition, the reduction in the number of properties is consistent with post-Fire regulations that stipulated that properties needed to be of a certain minimum size. Finally, our data end in 1675 (nine years after the Fire) so it could be the case that London had not reached a new stationary state – i.e., it is too early to tell if the number of properties converged to a new steady state. For these reasons, the main focus of our analysis is on quality as opposed to quantity. Nevertheless, in Appendix A.1, we examine the effect that the Fire had on the total number of properties and hearths in each parish.

1.4.1 Empirical strategy

To examine the effect of the Fire on the number of hearths per property, we use a DiD empirical strategy:

$$ln\left(Hearths_{ijt}\right) = \alpha_j + \delta PostFire_t + \beta Burned_j \times PostFire_t + \gamma' X_{jt} + \epsilon_{ijt} \tag{1.1}$$

$ln\left(Hearths_{ijt}\right)$ is the log number of hearths in property i in parish j in period t. The two periods are before the Fire and after the Fire. $Burned_j$ is an indicator variable that denotes whether property i was in a parish that experienced damage from the Fire. $PostFire_t$ is an indicator variable for the period after the Fire. X_{jt} is a vector of controls. Finally, α_j are parish fixed effects. We cluster the standard errors at the parish level. A null effect would suggest that there were no frictions to upgrading before the Fire – the quality of properties was optimal. By contrast, a positive effect suggests the presence of upgrading frictions which the Fire effectively removed.

For those interested in the cross-sectional regressions in each time period, the results are reported in Table A.7. In the pre-Fire period, the number of hearths per property in burned versus unburned parishes was statistically indistinguishable.

1.4.2 Results and discussion

Higher quality structures. Table 1.4 reports the impact that the Fire had on the number of hearths per property in the burned parishes relative to the unburned parishes. The estimate in column 1 shows that controlling for parish and time fixed effects, burned parishes saw a highly statistically significant increase of around 26.3% more hearths compared to unburned parishes. While in percentage terms this magnitude might seem large, given that the average number of hearths before the Fire was 3.83, this translates to an increase of 1.01 hearths.

Table 1.4: Effect of Fire on the number of hearths per property

	(1)	(2)	(3)	(4)
VARIABLES	ln(No. hearths)			
Parish Burned X Post Fire	0.263***	0.239**	0.219*	0.283**
	(0.092)	(0.098)	(0.127)	(0.116)
Observations	77,093	77,093	77,093	77,093
R-squared	0.009	0.010	0.010	0.014
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

There could be concerns that there are other time varying parish-level variables that are driving the results. For example, larger or richer parishes may recover faster from the Fire as they are able to bring together more resources. To address these concerns, in column 2, we include a series of parish controls interacted with $PostFire_t$. These include the number of properties in the parish before the Fire, the share of peers,[11] high-ranking military personnel (i.e., Colonel or Captain) and doctors living in the parish. The estimated effect remains robust to the inclusion of these time varying parish-level controls.

Next, to control for geographical characteristics, we classified the parishes into broader locations (i.e., abutting the City of London walls, within the walls and outside the walls). In column 3, we show that the results are stable to the inclusion of these broader locations-by-post fixed effects. Finally, we grouped parishes into terciles based on the number of hearths in the parish before the Fire. This is to control for the possibility that there may be persistence in the number of hearths – parishes with more hearths may rebuild with more hearths and those with fewer hearths may rebuild with fewer hearths. In column 4, we show that the results are relatively stable even when we include these pre-Fire hearth terciles-by-post fixed effects. Figure A.2 shows the binned scatter plot of the results in column 4.

Our results show that after the Fire, inhabitants of the parishes constructed more hearths per property. This suggests that there was indeed the presence of substantial frictions that was impeding development. By reducing the opportunity cost of waiting to rebuild to zero and forcing everyone to build at the same time, the Fire freed the parishes from the constraints imposed by their existing durable structures. This consequently spurred development through stronger cross-building spillovers and led to a new equilibrium which involved more hearths per property.

[11] These are Duke, Duchess, Marquess, Marchioness, Earl, Countess, Viscount, Viscountess, Baron, Baroness, Lord, Lady, Sir, Dame and Ambassador.

Finally, as our dependent variable has been log transformed, there could be issues of Jensen's inequality. In particular, running the regression with the log transformed dependent variable could result in an opposite treatment effect as compared to if we were to run the regression without taking logs. In Appendix A.2 we provide a discussion about this potential issue and show that we get a positive treatment effect in both the regression without logs and the regression in logs.

Effect varied with the level of damage. A priori, we should expect the effect of the Fire to vary with the level of damage. For example, in the extreme, if the Fire was so small that it only damaged one building, then the Fire would not have been effective in removing rebuilding frictions and there would be no widespread reconstruction.

We use two different approaches to examine such heterogeneous effects. First, we split the $Burned_j \times PostFire_t$ variable into two dummy variables – $SlightlyBurned_j \times PostFire_t$ and $CompletelyBurned_j \times PostFire_t$. As the names suggest, $SlightlyBurned_j$ refers to parishes where less than half of the parish (in terms of geographical area) was burned while the variable $CompletelyBurned_j$ refers to parishes where more than half of the parish was burned. Table 1.5 reports the results of this heterogeneous treatment effect regression. Across all columns, we see that the effect of the Fire was greater in parishes that were completely burned.

The second approach is to use whether the church in the parish was damaged as a proxy for the level of destruction in the parish due to the Fire. We think that this is reasonable given that the church was often the center of economic and social life during this period of time. To do this, we run regression 1.1 comparing burned parishes where the church was damaged to unburned parishes. In the same regression, we also compare burned parishes where the church was not damaged to unburned parishes. Table 1.6 reports the results. Across all columns, the effect of the Fire was greater in parishes where the church was also damaged. Both approaches suggest that the effect of the Fire was greater in parishes where the destruction was more widespread.

Table 1.5: Effect of the extent of Fire on the number of hearths per property

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Parish Completely Burned X Post Fire	0.403***	0.445***	0.471***	0.607***
	(0.064)	(0.089)	(0.073)	(0.102)
Parish Slightly Burned X Post Fire	0.173	0.136	0.144	0.192
	(0.117)	(0.133)	(0.146)	(0.130)
Observations	77,093	77,093	77,093	77,093
R-squared	0.012	0.013	0.013	0.018
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 1.6: Effect of the church being damaged on the number of hearths per property

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Parish burned and church damaged X Post Fire	0.393***	0.401***	0.400***	0.485***
	(0.073)	(0.100)	(0.115)	(0.122)
Parish burned but church not damaged X Post Fire	0.210*	0.189	0.190	0.252**
	(0.108)	(0.115)	(0.131)	(0.118)
Observations	77,093	77,093	77,093	77,093
R-squared	0.011	0.011	0.011	0.015
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Nonetheless, some of the positive effect that we find in Tables 1.4, 1.5 and 1.6 could be mechanical. This is because as noted by Field (2008), new houses had to be built according to strict regulations that specified the size and materials used. In addition, given the excess supply of land due to the plague, land was probably cheaper. This could lead to people wanting larger houses with more hearths per house. We discuss how we can rule out such mechanical effects in the next paragraph.

Effect varied with the level of damage in surrounding parishes. To rule out the mechanical effect of larger houses having more hearths, we split the $Burned_j \times PostFire_t$ variable into two dummy variables – $AllNeighorsBurned_j \times PostFire_t$ and $NotAllNeighorsBurned_j \times PostFire_t$ and re-run regression 1.1. If the increase in the number of hearths per property is purely due to larger houses, then it should not vary with the level of damage in the surrounding parishes.

Table 1.7: Effect of spatial spillovers on the number of hearths per property

	(1)	(2)	(3)	(4)
VARIABLES	ln(No. hearths)			
Not All Neighbors Burned X Post Fire	0.174	0.143	0.145	0.178
	(0.113)	(0.123)	(0.144)	(0.131)
All Neighbors Burned X Post Fire	0.395***	0.410***	0.409***	0.474***
	(0.069)	(0.098)	(0.102)	(0.124)
Observations	77,093	77,093	77,093	77,093
R-squared	0.010	0.011	0.011	0.014
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 1.7 shows that spatial spillovers matter. Burned parishes that were completely surrounded by other burned parishes experienced building investments that were two to three times higher than burned parishes that were not completely surrounded by burned parishes. This regression shows us strong evidence that the increase in the number of hearths per building is driven by cross-building spillovers and not the mechanical effect of larger properties.

1.4.3 Robustness checks

Dropping parishes which merged after the Fire. One concern could be that the Fire led to the merging of some parishes and so our results could be driven by these enlarged parishes which might have more resources. In Table A.8 we re-run regression 1.1 using only parishes that did not merge with other parishes after the Fire. Reassuringly, the coefficient estimates remain very similar to our baseline results.

Using different control groups. There could be concerns that some of the unburned parishes in our control group may not be appropriate for our analysis. To see this, consider a hypothetical unburned parish (parish U) that is surrounded by many burned parishes. Given the destructive nature of the Fire, it is somewhat surprising that parish U did not suffer any damage from the Fire. This could suggest that parish U is fundamentally different from its neighboring parishes that were burned. For example, parish U could have been more wealthy and hence more able to quickly mobilize fire-fighting efforts. It could also be the case that more buildings in parish U were made of bricks as opposed to wood. In addition, parish U could have also pre-empted the spread of the Fire by tearing down buildings that were near to the parishes that were burning. The Fire Court records suggest that this indeed happened in parishes such as St Botolph Bishopsgate and St Mary-le-Strand. Therefore, it might not be suitable to use such parishes as the control group to the burned parishes.

To address this concern, we run separate regressions based on two samples. First, a "nearby" sample which consists of all burned parishes plus unburned parishes that share a boundary with any burned parishes. The results for this are reported in Table A.9. Second, a "further away" sample which consists of all burned parishes and unburned parishes that do not share any boundaries with any burned parishes. The results using this sample are reported in Table A.10. The results using both samples are very similar, suggesting that our analysis is not sensitive to the choice of control groups.

Accounting for zeros in the outcome variable. There are 2,637 observations which are recorded as having zero hearths in the property. Taking logs results in these observations dropping out of the regression. Therefore, to account for the zeros in the outcome variables, we adopt two approaches. First, applying the inverse hyperbolic sine transform to hearths. Second, using a Poisson pseudo-likelihood (PPML) regression. The results are reported in Tables A.11 and A.12 respectively. While the estimated effects remain positive, the magnitudes are now halved and are imprecisely estimated.

Trimming extreme values of the outcome variable. As another robustness check, we drop the top and bottom one percentile of $ln\left(Hearths_{ijt}\right)$. While the estimated effects are now halved, they remain positive and are mostly statistically significant. The results of this robustness check are reported in Table A.13.

Checking for parallel trends. A key assumption of a DiD strategy is that of parallel trends. There are two methods to argue that the number of hearths per property in the burned versus unburned parishes had the same trends before the Fire. The first method is to rely on the historical context. The historical setting suggests that the Fire spread based on where the wind blew and not due to the economic or social characteristics of the parish. While the Fire started in Pudding Lane which was in the eastern part of the City of London, contemporaneous reporting by the The London Gazette (1666) notes that due to the "violent Easterly wind", the Fire spread mostly to the west. As a result, almost all the parishes that were damaged were to the west of Pudding Lane.

The wind blowing from the east to the west during the Fire is an important point. This is because Heblich et al. (2020) show that in England, the wind usually blows from the west or south-west. Therefore, we can make an argument that whether a parish ended up being burned was unexpected, random and independent of pre-trends.

The second method would be to run a placebo DiD regression to compare the number of hearths per property in the burned versus unburned parishes in the periods before the Fire. We should find no effect if there are parallel trends. Unfortunately, due to data limitations, we are not able to do so in the most robust manner. This is because the hearth tax was introduced in 1662 and for the pre-Fire period, we only have data from the 1662, 1664, 1665 and 1666 hearth tax records. Due to the differences in the scope and range of the pre-Fire hearth tax assessments, almost all parishes (65 out of 70) appear only in one year. In addition, Table 1.8 shows that in some years, the data we have were either all for burned or unburned parishes. Therefore, we had to pool all the 1662, 1664, 1665 and 1666 data to form the pre-Fire period in our regressions.

Table 1.8: Pre-Fire data

Year	Burned parishes in the data	Unburned parishes in the data	Total
1662	16	0	16
1664	0	2	2
1665	0	2	2
1666	38	17	55

To run placebo regressions to test for pre-trends, we would need data for both burned and unburned parishes in at least two pre-Fire years. However, with the exception of 1666, the other three pre-Fire years only consist of data from either burned or unburned parishes. In order to try our best to provide statistical evidence to rule out pre-trends, what we can do is to classify the pre-Fire period into two categories. First, the year 1666 would be period $t-1$ and the years 1662 and 1664 would be $t-2$.[12] This method has some limitations such as assuming that the data in 1664 are very similar to 1662 and the two unburned parishes in 1664 are representative of all the other unburned parishes. Nevertheless, accepting these limitations allows us to run the following placebo regression to test for pre-trends:

$$ln\left(Hearths_{ijt}\right) = \alpha_j + \delta PlaceboFire_t + \beta Burned_j \times PlaceboFire_t + \gamma' X_{jt} + \epsilon_{jt}$$

$PlaceboFire_t$ is an indicator variable for the period after a placebo fire. We set this as the period after 1665. Finding a large and statistically significant effect from this phantom event would cast serious doubts on the validity of our identification strategy. If our regression passes the parallel trends test, then we should expect β to be small and statistically insignificant. Table 1.9 shows that this is indeed what we get when we run this placebo regression, suggesting the absence of pre-trends.

Table 1.9: Effect of placebo Fire on the number of hearths per property

VARIABLES	(1) ln(No. Hearths)
Parish Burned X Post Placebo Fire	-0.009
	(0.023)
Observations	40,517
R-squared	0.004
Parish FE	✓
Post FE	✓

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

[12] We do not include year 1665 because the two parishes that appear in 1665 are south of the river and are hence very different from the other parishes in the data for the placebo regression.

1.5 The Effect of Pragmatic Legal Rulings

1.5.1 Overview

In the previous section, we found that the 1666 Great London Fire resulted in a higher number of hearths per property in burned parishes relative to unburned parishes. While our results suggest that individuals had positive expectations about how much other individuals in their parish would be investing, it does not tell us what is driving these expectations. Therefore, in this section, we examine if legal rulings could be a driver of these expectations.

1.5.2 Defining pragmatic legal rulings

The Fire Court judges were given the power to completely void existing contracts or alter the terms of these contracts. Voiding the existing contract means that both the landlord and tenant surrender their contracts. This allows both parties to negotiate a new contract with each other or other parties. In addition, voiding the contract does away with the judging panels arbitrarily setting a new rent and lease. By contrast, altering the terms of an existing contract means that the tenant remains the same but the Fire Court decrees a new rent and/or length of lease. In addition, the Fire Court would also decree that either the landlord or tenant rebuilds, or that both parties are to contribute towards the rebuilding. In our paper, we define the voiding of existing contracts and the assigning of the cost of rebuilding to the landlord as pragmatic legal rulings. 12.7% of cases fall into this category.

We define such rulings as pragmatic because they help to facilitate a higher number of hearths per property. First, tenants are likely to be more credit-constrained compared to landlords and are hence more likely to rebuild at a lower quality (i.e., fewer hearths per property). Second, assigning the rebuilding to the landlord represents a clear assignment and alignment of property-rights. Third, since the occupant is responsible for paying the hearth tax, if the tenant was assigned the rebuilding, she is likely to rebuild with fewer hearths to reduce her tax burden. For these reasons, assigning the rebuilding to the landlord rather than the tenant facilitates the rebuilding of London.

In theory, we could have expanded our definition of pragmatic rulings to also include cases where the judges altered existing contracts (i.e., did not void the contract) but assigned the rebuilding responsibility to the landlord. However, in such cases the Fire Court's rulings were often multi-dimensional. For example, it could be the case that although the rebuilding responsibility was assigned to the landlord, the judges could have decreed a lower rent. In this instance, the landlord may then choose to rebuild at a lower quality since the rent she is receiving is now lower. In order to circumvent the issue of multi-dimensional rulings, we focus on the most extreme of case outcomes – cases where the Fire Court voided existing contracts and assigned the rebuilding responsibility to the landlord.[13]

1.5.3 Model: Legal rulings, expectations and investment

How exactly did legal rulings drive expectations and hence help to coordinate investment (i.e., the number of hearths in each property)? We show this using a Nash (1950) bargaining game where tenants and landlords bargained over the terms of rebuilding. The bargaining game consists of two stages. In Stage 1, in each parish j, the landlord and tenant of each property i bargain over a contract given their respective outside options. The outside options are based on the rulings established by the Fire Court in its initial cases for each parish. Therefore, the outside options vary across parishes. For simplicity, we suppress the subscripts j and i. We define the contract $\{r, t, I^l\}$ in terms of the annual rent (r), the tenancy length (t), and the amount of contributions (investment) that the landlord makes towards the rebuilding (I^l).[14] If the tenant and landlord reach an agreement, they move to the second stage where the tenant decides on her amount of contributions (investment) towards the rebuilding. The total amount of building investment (measured in terms of the number

[13]Finally, we could have also expanded our definition of pragmatic rulings to include cases where the judges voided the existing contracts but decreed cost sharing in the rebuilding. However, the reason why we do not do this is because we wanted our definition to reflect the complete burden of rebuilding falling on the landlord. This will be clearer when we discuss the model in the next section.

[14]More accurately, the parties bargain over the split of the total surplus. In doing so, the parties are implicitly choosing $\{r, t, I^l\}$.

of hearths in the property) is given by the sum of the landlord's investment (determined in the first stage) and the tenant's investment (determined in the second stage). If they fail to reach an agreement, they bring their case to the Fire Court. In this framework, the Fire Court's rulings affect the outside options and hence the bargaining dynamics between the landlords and tenants. The model is solved by backward induction.

Solving the Nash bargaining game: Stage 2

The tenant's problem in Stage 2 is to choose I^t to maximize utility given the contract $\left\{r, t, I^l\right\}$ that was determined in Stage 1:

$$\underset{I^t}{max}\ U\left(r, t, I^l, I^t\right) = \frac{1-\beta^t}{1-\beta}\left[h(I^l, I^t) - r\right] + \frac{\beta^t}{1-\beta}u^0 - pI^t \tag{1.2}$$

β is the discount factor. p is cost per unit of investment. u^0 is the tenant's utility after the tenancy ends. We assume that $h(I^l, I^t)$ is concave and that the amount of investments that the landlord and tenant make towards the building are complements ($\frac{\partial h}{\partial I^l \partial I^t} > 0$).

The first order condition is:

$$\frac{1-\beta^t}{1-\beta}h'(I^l, I^t) = p \tag{1.3}$$

Equation 1.3 suggests that the tenant's investment does not depend on the rent.

Proposition 1. The tenant's investment is increasing in the tenancy length.

Proof: From equation 1.3, let $\psi\left(t, I^l, I^t\right) \equiv \frac{1-\beta^t}{1-\beta}h'(I^l, I^t) - p$. By the implicit function theorem and since $h(I^l, I^t)$ is concave:

$$\begin{aligned}\frac{\partial I^t}{\partial t} &= -\frac{\psi_t}{\psi_{I^t}} \\ &= \frac{-\beta^t \beta h'(I^l, I^t)}{-\frac{1-\beta^t}{1-\beta} h''(I^l, I^t)} > 0 \ \square \end{aligned} \tag{1.4}$$

Let the optimal tenant investment be denoted as: $I^{t*} \equiv g\left(t, I^l\right)$. Therefore, total investment is: $I\left(t\right) \equiv g\left(t, I^l\right) + I^l$

Solving the Nash bargaining game: Stage 1

In this stage, the tenant and landlord bargain over the surplus given their respective outside options π^c and u^c. Their outside options are based on the Fire Court's rulings in the initial cases. We assume that π^c and u^c vary across parishes. We represent the distribution of Fire Court decisions in the initial cases as $F(r, t, I^l)$. λ is the bargaining weight which we assume to be exogenous. The Nash bargaining game solution can be characterized as:

$$\underset{\left\{r,t,I^l\right\}}{max} \quad \left[\Pi\left(r, t, I^l\right) - \pi^c\right]^{\lambda} \left[U\left(r, t, I^l\right) - u^c\right]^{(1-\lambda)} \tag{1.5}$$

where the landlord's utility is $\Pi\left(r, t, I^l\right) = \frac{1-\beta^t}{1-\beta} r + \frac{\beta^t}{1-\beta} r^0\left(I(t)\right) - pI^l$. $r^0\left(I(t)\right)$ is the rent that the landlord receives from the next tenant after the tenancy agreement with the current tenant expires. In addition,

$$\pi^c = \int\int\int \frac{1-\beta^y}{1-\beta} x + \frac{\beta^y}{1-\beta} r^0\left(I(y)\right) - pz \; dF(x, y, z)$$

and

$$u^c = \int\int\int \frac{1-\beta^y}{1-\beta}\left[h(I(y)) - x\right] + \frac{\beta^y}{1-\beta} u^0 - pg(y, z) \; dF(x, y, z)$$

The Nash bargaining solution for the landlord is:

$$\Pi\left(r,t,I^{l}\right) = \lambda[\Pi\left(r,t,I^{l}\right) + U\left(r,t,I^{l}\right) - \pi^{c} - u^{c}] + \pi^{c} \tag{1.6}$$

and that for the tenant is:

$$U\left(r,t,I^{l}\right) = (1-\lambda)[\Pi\left(r,t,I^{l}\right) + U\left(r,t,I^{l}\right) - \pi^{c} - u^{c}] + u^{c}$$

Rearranging equation 1.6, we get that:

$$\begin{aligned}(1-\lambda)\left[\frac{1-\beta^{t}}{1-\beta}r + \frac{\beta^{t}}{1-\beta}r^{0}\left(I(t)\right) - pI^{l}\right] - \pi^{c} \\ = \lambda\left[\frac{1-\beta^{t}}{1-\beta}\left[h(I(t)) - r\right] + \frac{\beta^{t}}{1-\beta}u^{0} - pg(I^{l},t) - Q^{c}\right]\end{aligned} \tag{1.7}$$

where $Q^{c} \equiv \pi^{c} + u^{c} = \int\int\int \frac{1-\beta^{y}}{1-\beta}h(I(y)) + \frac{\beta^{y}}{1-\beta}\left[r^{0}\left(I(y)\right) + u^{0}\right] - pI(y)\; dF(x,y,z)$

Next, we assume that the judging panel's preferences for the landlord's contribution to the rebuilding is orthogonal to r and t. In other words, $F(x,y,z) \equiv F_{XY}(x,y)F_{Z}(z)$. Assuming that $F(x,y,z) \equiv F_{XY}(x,y)F_{Z}(z)$ has two implications. First, this assumption implies that the sum of the outside options is not affected by the transfer of the burden to rebuild:

$$Q^{c} = \int\int \frac{1-\beta^{y}}{1-\beta}h(I(y)) + \frac{\beta^{y}}{1-\beta}\left[r^{0}\left(I(y)\right) + u^{0}\right] - pI(y)\; dF(x,y)$$

Second, while the sum of the outside options is not affected by the transfer of the burden to rebuild, the outside option of the landlord still depends on $F_{Z}(z)$. Given these two implications, equation 1.7 can be expressed as:

$$
\begin{aligned}
&(1-\lambda)\left[\frac{1-\beta^t}{1-\beta}r+\frac{\beta^t}{1-\beta}r^0\left(I(t)\right)-pI^l\right]-\pi^c\left(F_Z(z)\right)\\
&=\lambda\left[\frac{1-\beta^t}{1-\beta}\left[h(I(t))-r\right]+\frac{\beta^t}{1-\beta}u^0-pg(I^l,t)-Q^c\right]
\end{aligned}
$$

Now suppose that there is a contract $\left\{\bar{r},\bar{t},\bar{I}^l\right\}$ that satisfies the Nash bargaining game solution. However, we switch from $F_Z(z)$ to $F'_Z(z)$, where $F'_Z(z)$ first order stochastically dominates $F_Z(z)$. Recall that $F(r,t,I^l)$ represents the distribution of Fire Court rulings in the initial cases. In our empirical context, moving from $F_Z(z)$ to $F'_Z(z)$ corresponds to the initial cases getting assigned judging panels that have a greater probability of voiding the existing contracts and assigning the cost of rebuilding to the landlord. As explained in the previous section, this is what we define as pragmatic legal rulings.

Proposition 2. The landlord's outside option (π^c) *falls when the initial cases are assigned judging panels that have a greater preference for the landlord to contribute more to the rebuilding.*

Proof: Since $F'_Z(z)$ first order stochastically dominates $F_Z(z)$, this implies that the landlord's outside option under $F'_Z(z)$ is now smaller:

$$
\begin{aligned}
&F'_Z(z)\leq F_Z(z)\ \forall z \quad \text{and} \quad F'_Z(z)<F_Z(z)\text{ for some } z \qquad (1.8)\\
&\Rightarrow \pi^c\left(F'_Z(z)\right)<\pi^c\left(F_Z(z)\right)
\end{aligned}
$$

The last inequality is because $\pi^c=\int\int\frac{1-\beta^y}{1-\beta}x+\frac{\beta^y}{1-\beta}r^0\left(I(y)\right)\ dF_{XY}(x,y)-p\int z\ dF_Z(z)$. Since $F'_Z(z)<F_Z(z)$, $p\int z\ dF'_Z(z)>p\int z\ dF_Z(z)$ and so $\pi^c\left(F'_Z(z)\right)<\pi^c\left(F_Z(z)\right)$.$\square$

Since the landlord now has a smaller outside option, the contract $\left\{\bar{r},\bar{t},\bar{I}^l\right\}$ no longer satisfies the Nash bargaining game solution and equation 1.7 no longer holds with equality. Instead, the landlord now has too much of the surplus and so:

$$(1-\lambda)\left[\frac{1-\beta^{\bar{t}}}{1-\beta}\bar{r}+\frac{\beta^{\bar{t}}}{1-\beta}r^{0}\left(I(\bar{t})\right)-p\bar{I}^{l}\right]-\pi^{c}\left(F_{Z}'(z)\right)$$
$$>\lambda\left[\frac{1-\beta^{\bar{t}}}{1-\beta}\left[h(I(\bar{t}))-\bar{r}\right]+\frac{\beta^{t}}{1-\beta}u^{0}-pg(\bar{I}^{l},\bar{t})-\bar{Q}^{c}\right] \tag{1.9}$$

In order to achieve equality, changes in the Nash bargained contract should (1) lower the left-hand side of the inequality and increase the right-hand side or (2) increase the right-hand side more than the left-hand side.

Proposition 3. The Nash bargained rent $(\bar{r})$ *decreases when the judging panels have a greater preference for the landlord to contribute more to the rebuilding.*

Proof: Referring to inequality 1.9, since $\frac{\partial LHS}{\partial \bar{r}}=\frac{1-\beta^{\bar{t}}}{1-\beta}(1-\lambda)>0$ and $\frac{\partial RHS}{\partial \bar{r}}=-\frac{1-\beta^{\bar{t}}}{1-\beta}\lambda<0$, in order for the left-hand side to equal to the right-hand side, $\bar{r}$ has to decrease. A decrease in $\bar{r}$ decreases the left-hand side and increases the right-hand side.$\square$

Proposition 4. The Nash bargained landlord investment to the rebuilding $(\bar{I}^{l})$ *increases when the judging panels have a greater preference for the landlord to contribute more to the rebuilding.*

Proof: Referring to inequality 1.9, since $\frac{\partial LHS}{\partial \bar{I}^{l}}=-p(1-\lambda)<0$ and $\frac{\partial RHS}{\partial \bar{I}^{l}}=-p\lambda g_{\bar{I}^{l}}\left(\bar{I}^{l},\bar{t}\right)=p\lambda>0$, in order for the left-hand side to equal to the right-hand side, $\bar{I}^{l}$ has to increase. An increase in $\bar{I}^{l}$ decreases the left-hand side and increases the right-hand side.$\square$

Proposition 5. If the relative bargaining weight of the landlord is more than the relative marginal benefit of increasing the tenancy length, then the Nash bargained tenancy length $(\bar{t})$ *increases when the judging panels have a greater preference for the landlord to contribute more to the rebuilding.*

Proof: Referring to inequality 1.9, $\frac{\partial LHS}{\partial \bar{t}}=(1-\lambda)\left\{(-ln\beta)\beta^{\bar{t}}\left[\bar{r}-r^{0}(I(\bar{t}))\right]+\frac{\beta^{\bar{t}}}{1-\beta}r_{1}^{0}(I(\bar{t}))g_{\bar{t}}\left(\bar{I}^{l},\bar{t}\right)\right\}$ $=(1-\lambda)\Pi_{\bar{t}}\left(\bar{r},\bar{t},\bar{I}^{l}\right)$ and $\frac{\partial RHS}{\partial \bar{t}}=\lambda\left\{(-ln\beta)\beta^{t}\left[h(I(\bar{t}))-\bar{r}-u^{0}\right]+\left[\frac{\beta^{\bar{t}}}{1-\beta}h'(I(\bar{t}))-p\right]g_{\bar{t}}\left(\bar{I}^{l},\bar{t}\right)\right\}=$

$\lambda U_{\bar{t}}\left(\bar{r}, \bar{t}, \bar{I}^l\right)$. To achieve equality, we need:

$$\begin{aligned} \frac{\partial LHS}{\partial \bar{t}} &< \frac{\partial RHS}{\partial \bar{t}} \\ \Rightarrow (1-\lambda)\Pi_{\bar{t}}\left(\bar{r}, \bar{t}, \bar{I}^l\right) &< \lambda U_{\bar{t}}\left(\bar{r}, \bar{t}, \bar{I}^l\right) \\ \Rightarrow \frac{\lambda}{1-\lambda} &> \frac{\Pi_{\bar{t}}\left(\bar{r}, \bar{t}, \bar{I}^l\right)}{U_{\bar{t}}\left(\bar{r}, \bar{t}, \bar{I}^l\right)} \quad \square \end{aligned}$$

This gives us a necessary and sufficient condition for $\bar{t}$ to increase. In other words, if the relative bargaining weight of the landlord is more than the relative marginal benefit of increasing the tenancy length, then $\bar{t}$ increases. Given the historical context that tenants are obliged to repair or rebuild the premises in the event of disasters or wars, this condition is likely to hold. In addition, in the Fire Court data, we see that the judging panels decreed that the tenant had to rebuild 70.9% of the time.

Putting everything together, our model suggests that if the initial cases are assigned judging panels that have a greater preference for the landlord to contribute more to the rebuilding, then this lowers the landlord's outside option (proposition 2). As a result of the lowering of the landlord's outside option, the Nash bargained annual rent (r) decreases (proposition 3), the amount of investment that the landlord makes towards the rebuilding (I^l) increases (proposition 4), and the effect on the tenancy length (t) increases (proposition 5). Crucially, our model shows us that by changing outside options, the rulings of the Fire Court affected **all** tenants and landlords even if they did not bring their case to the Fire Court. This is how legal rulings affect expectations.

Empirical implication

In our empirical analysis, we estimate the change in the average number of hearths per property in parish j as a result of the initial cases in the parish getting assigned judging panels that have a greater propensity to void existing contracts and assign the rebuilding to the landlord (i.e., pragmatic rulings). This corresponds loosely to:

$$
\begin{aligned}
\frac{\partial E_j(I_i)}{\partial E_j(F_Z(z))} &= \frac{\partial E_j(I_i^t)}{\partial E_j(F_Z(z))} + \frac{\partial E_j(I_i^l)}{\partial E_j(F_Z(z))} \\
&= \underbrace{\frac{\partial E_j(I_i^t)}{\partial E_j(r_i)}}_{=0 \text{ by eq. } 2} \times \underbrace{\frac{\partial E_j(r_i)}{\partial E_j(F_{Z_j}(z))}}_{<0 \text{ by prop. } 3} + \underbrace{\frac{\partial E_j(I_i^t)}{\partial E_j(I_i^l)}}_{>0 \text{ by assump.}} \times \underbrace{\frac{\partial E_j(I_i^l)}{\partial E_j(F_{Z_j}(z))}}_{>0 \text{ by prop. } 4} + \underbrace{\frac{\partial E_j(I_i^t)}{\partial E_j(t_i)}}_{>0 \text{ by prop. } 1} \times \underbrace{\frac{\partial E_j(t_i)}{\partial E_j(F_{Z_j}(z))}}_{>0 \text{ by prop. } 5} \\
&+ \underbrace{\frac{\partial E_j(I_i^l)}{\partial E_j(r_i)}}_{\text{ambiguous}} \times \underbrace{\frac{\partial E_j(r_i)}{\partial E_j(F_{Z_j}(z))}}_{<0 \text{ by prop. } 3} + \underbrace{\frac{\partial E_j(I_i^l)}{\partial E_j(I_i^l)}}_{=1} \times \underbrace{\frac{\partial E_j(I_i^l)}{\partial E_j(F_{Z_j}(z))}}_{>0 \text{ by prop. } 4} + \underbrace{\frac{\partial E_j(I_i^l)}{\partial E_j(t_i)}}_{\text{ambiguous}} \times \underbrace{\frac{\partial E_j(t_i)}{\partial E_j(F_{Z_j}(z))}}_{>0 \text{ by prop. } 5}
\end{aligned} \tag{1.10}
$$

The second line gives us the effect on the tenant's investment and this effect is unambiguously positive. However, the effect on the landlord's investment (third line) is ambiguous. The signs of $\frac{\partial E_j(I_i^l)}{\partial E_j(r_i)}$ and $\frac{\partial E_j(I_i^l)}{\partial E_j(t_i)}$ are ambiguous because the landlord can in principle trade off a higher amount of investment to the building process with a lower rent or longer tenancy length. This happens because there are three variables $\{r, t, I^l\}$ that are governed by a single Nash bargaining equation (see equation 1.6). If the positive effect on the tenant's investment (second line) dominates the ambiguous effect on the landlord's investment (third line) then pragmatic legal rulings can result in a higher number of hearths per property.

To conclude this section, our model shows that even though landlords and tenants of different properties are bargaining separately and do not bring their case to the Fire Court, they end up choosing similar levels of hearths per property. This is because they have the same focal point and hence expectations of what others will do. This focal point is how the Fire Court ruled in the initial cases in their parish.

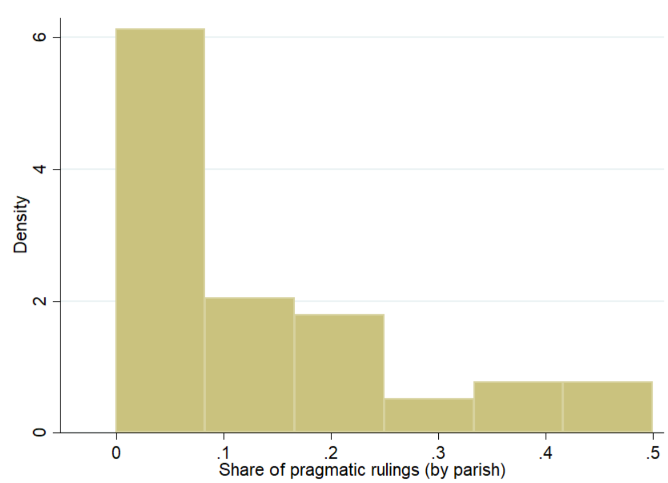
1.5.4 Empirical strategy

To examine the effect of legal rulings, we continue to use a DiD empirical strategy:

$$ln\left(Hearths_{ijt}\right) = \alpha_j + \delta PostFire_t + \beta PragmaticRulings_j \times PostFire_t + \gamma' X_{jt} + \epsilon_{ijt} \quad (1.11)$$

$ln\left(Hearths_{ijt}\right)$ is the log number of hearths in property i in parish j in period t. The two periods are before the Fire and after the Fire. $PragmaticRulings_j$ denotes the share of initial cases in parish j where the Fire Court judging panels' rulings were pragmatic.[15] Specifically, this is the share of cases in parish j where the Fire Court judging panels voided the existing contracts and assigned the rebuilding to the landlord. Figure 1.2 shows the distribution of the share of pragmatic rulings across the parishes. $PostFire_t$ is an indicator variable for the period after the Fire. X_{jt} is a vector of controls. Finally, α_j are parish fixed effects. We cluster the standard errors at the parish level.

Figure 1.2: Distribution of the share of pragmatic rulings across parishes



[15] As we only have data from the first four out of nine volumes of the Fire Court cases, these figures are calculated based on the first four volumes.

It is important to note that we are not able to distinguish in the data properties that went to the Fire Court and those that did not. The regression therefore includes all properties in the parish – those that went to the Fire Court and those that did not. However, this should not affect our results substantially since the proportion of properties in each parish that went to the Fire Court is a relatively small number. Based on the initial cases, the average proportion of properties in each parish that went to the Fire Court was 6%, the median was 4% and the maximum was 30%. In Table A.6 we report the proportion of properties in each parish where the landlord and the tenant went to the Fire Court (based on the data that we have). Therefore, in the regression, β also tells us whether the rulings in a small share of properties in the parish that went to the Fire Court affected the quality of other properties in the parish.

For those interested in the cross-sectional regressions in each time period, the results are reported in Table A.14. In the pre-Fire period, the number of hearths per property in parishes where all the legal rulings were pragmatic versus parishes with no pragmatic legal rulings was statistically indistinguishable.

Recall that in the previous section, our DiD regression compares burned parishes to unburned parishes. As a result, there could be concerns that any positive effect is purely mechanical since rebuilt properties had to be built according to strict regulations that specified the size and materials used. However, in this section, since our sample consists only of burned parishes, the DiD strategy helps to net off these mechanical effects. This allows us to more cleanly attribute the effect that we estimate to legal rulings.

1.5.5 Results and discussion

Higher quality structures. Table 1.10 reports whether the rulings in a small share of properties in the parish that went to the Fire Court affected the quality of other properties in the parish (i.e., number of hearths per property). The estimate in column 1 shows that controlling for parish and year fixed effects, parishes where all the initial cases saw pragmatic Fire Court rulings experienced a highly statistically significant increase of around 144% more hearths compared to parishes where all the initial cases saw unpragmatic rulings.

Importantly, the share of pragmatic rulings differs across parishes because of both exogenous and endogenous reasons. Therefore, in our regression, we control for as many endogenous reasons as we can. In column 2, we include a series of parish controls interacted with $PostFire_t$. These include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. This helps to address concerns that our results could be driven by the politics and resources of the parishes. Reassuring, our results remain extremely stable to the inclusion of these controls.

In column 3, we show that the results are stable to the inclusion of broader locations-by-post fixed effects. Finally, in column 4, we include pre-Fire hearth terciles-by-post fixed effects. The estimated effect continues to be robust to the inclusion of these controls. In particular, parishes where all the initial cases saw pragmatic Fire Court rulings experienced a highly statistically significant increase of around 98.1% more hearths compared to parishes where all the initial cases resulted in unpragmatic rulings. Given that the average number of hearths before the Fire was 3.83, this translates to an increase of 3.76 hearths. Expressed in a different way, what our result suggests is that in terms of the share of pragmatic rulings, going from the 25th percentile parish (0% pragmatic rulings) to the 75th percentile parish (20% pragmatic rulings) resulted in a 19.6% increase in the number of hearths. In absolute terms, this corresponds to an increase of 0.75 hearths. Figure A.3 shows the binned scatter plot of the results in column 4.

Table 1.10: Effect of legal rulings on the number of hearths per property

	(1)	(2)	(3)	(4)
VARIABLES	ln(No. hearths)			
Pragmatic X Post Fire	1.444***	1.253***	1.103***	0.981***
	(0.449)	(0.393)	(0.286)	(0.246)
Observations	31,582	31,582	31,582	31,582
R-squared	0.014	0.024	0.026	0.031
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

The results show us that while only a small share of properties in each parish went to the Fire Court, the rulings of these few cases had an outsized effect on the quality of other buildings in the parish. Why would this be the case? We argue that this is because the small share of cases was enough to anchor expectations. This positive result is even more remarkable considering that the hearth tax was introduced in 1662. Occupants of properties would have been incentivized to rebuild with fewer hearths to avoid the tax. Despite this, we still see a positive effect due to legal rulings. This suggests the possibility that even before the Fire, there was latent demand for structures with more hearths. Consequently, the simultaneous building after the Fire led to greater cross-building spillovers and helped to address this demand. Our results hence provide us with evidence that pragmatic legal rulings can indeed anchor expectations of what others will do.

Finally, our regression uses parish-level variations in Fire Court rulings. A natural question to ask is why do the Court rulings in your own parish matter? We take two approaches to address this. First, using the historical context. In early modern London, most interactions took place at the parish-level. Individuals often worked, lived and worshiped in the same parish. Moreover, parishes were given quite a bit of autonomy in civil matters. For example, the Highways Act 1555 made road maintenance the responsibility of the parish and Poor Relief Act 1601 (Poor Law) outlined the responsibility of the parish to look after its own poor. Therefore, because of the context, we argue that what is most salient to inhabitants of the parish is what happens within their own parish.

Second, we show statistical evidence that the rulings of previous cases in your own parish predicts future rulings in your parish. To show this, we run the following regression using the Fire Court cases of parishes that appear in the hearth tax data:

$$PragmaticRuling_{ijp} = \theta_p + \beta PragmaticRulingFirstFewCases_j + \lambda' X_{ijp} + \epsilon_{ijp}$$

$PragmaticRuling_{ijp}$ is a dummy variable that indicates whether the judging panel p for case i in parish j decreed a pragmatic ruling. $PragmaticRulingFirstFewCases_j$ is the share of pragmatic rulings in the first few cases preceding the current case in parish j. When running the regressions, we try different definitions of "first few cases". For example, the first two cases, the first three cases, etc. Taking the average across the first few cases accounts for the fact that the first case may not be precedential and precedents may take some time to be firmly established. θ_p are judging panel fixed effects. X_{ijp} is a vector of controls. These include pre-Fire case characteristics such as the degree of subletting in the property, the number of years left in the tenancy, the rent, the fine paid to secure the contract and whether the tenant spent any money to improve the property. Importantly, the vector of controls also includes the share of pragmatic rulings in other parishes before case i in parish j. The standard errors are clustered at the parish level.

Table 1.11 reports the results. In column 1, the definition of first few cases is the first case, in column 2, the definition of the first few cases is the first two cases, and so on in the other columns.

Across all columns, the coefficient estimate of β is positive. This suggests that past rulings in your own parish predicts future rulings. In addition, the coefficient estimates increase as we move across the columns. This reflects the fact that the first case may not be precedential and precedents may take some time to be firmly established. Column 5 suggests that if the first five cases in your parish had pragmatic rulings, the probability that the current case receives a pragmatic ruling increases by 40.5%-points. Therefore, the Court rulings in your own parish matter because they predict future rulings in your parish.

Table 1.11: Effect of past rulings in your own parish on current ruling

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Whether current case ruling is pragmatic				
Share of pragmatic rulings (first few cases)	0.024	0.022	0.202*	0.327**	0.405*
	(0.061)	(0.075)	(0.117)	(0.144)	(0.199)
Observations	303	246	195	163	139
R-squared	0.350	0.417	0.443	0.510	0.488
Judging panel FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
First few cases	1	2	3	4	5

Notes: Case controls include pre-Fire case characteristics such as the degree of subletting in the property, the number of years left in the tenancy, the rent, the fine paid to secure the contract and whether the tenant spent any money to improve the property. It also includes the share of pragmatic rulings in other parishes before the current case. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Effect holds when controlling for spillovers from neighboring parishes. Next, we check if our results are sensitive to potential spatial spillovers. In particular, burned parishes with Fire Court cases tend to be located near each other. Therefore, a burned parish with Fire Court cases not only generated spillovers to other burned parishes but also received inward spillovers from these other parishes. If these spillovers are large, our estimated effects of legal rulings within each parish could be overstated. Therefore, we include as a control the weighted share of cases in neighboring burned parishes where the Fire Court judging panels decreed pragmatic rulings. Table 1.12 shows that our results are robust to controlling for the legal outcomes in neighboring parishes.

Table 1.12: Effect of legal rulings on the number of hearths per property (controlling for spillovers)

	(1)	(2)	(3)	(4)
VARIABLES	ln(No. hearths)			
Pragmatic X Post Fire	1.485***	1.149***	1.119***	0.963***
	(0.478)	(0.326)	(0.278)	(0.253)
Pragmatic Spillover X Post Fire	-0.513	0.230	0.013	0.349
	(0.635)	(0.358)	(0.315)	(0.286)
Observations	31,582	31,582	31,582	31,582
R-squared	0.015	0.025	0.026	0.031
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. We also include the log number of cases in neighboring parishes. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Adding in other dimensions of the rulings as controls. The Fire Court judges were given the power to decree who rebuilds, the rent, as well as length of the new contract. Such multi-dimensional rulings make it difficult to define what constitutes pragmatic rulings that helped to facilitate the rebuilding of London. In order to circumvent the issue of multi-dimensional rulings, in our analysis, we had focused on the most extreme of pragmatic case outcomes – cases where the Fire Court voided existing contracts and assigned the rebuilding to the landlord. Nonetheless, it could well be the case that the newly decreed rent or tenancy length could be playing a role in anchoring expectations and consequently, the number of hearths per property in each parish.

To address this concern, we include as controls the average change in rent and tenancy length in each parish arising from the Fire Court rulings. For example, if there were five cases in parish A that went to the Fire Court, and the judging panel increased the tenancy length for all five of the cases by 10 years, then the average change in tenancy length arising from the Fire Court rulings for parish A would be 10 years. Referring to Table A.15, the coefficient estimate on our treatment variable $PragmaticRulings_j \times PostFire_t$ remains stable to the inclusion of the other dimensions of the Fire Court's rulings. The coefficient estimates on the average change in the tenancy length interacted with post are extremely close to zero and statistically insignificant. The coefficient estimates on the average change in the rent interacted with post while significant, is relatively small in magnitude. Taken together, the results across the columns provide evidence that it is indeed pragmatic Fire Court rulings (as defined by the share of initial cases where the judging panels voided the contracts and assigned the rebuilding to the landlord) that are affecting individuals' expectations and not the other dimensions of the Fire Court's decisions.

1.5.6 Competing hypotheses/mechanisms

The results in Table 1.10 show us that while only a small share of properties in each parish went to the Fire Court, the rulings of these few cases had an outsized effect on the quality of other buildings in the parish. We argue that this is because the small share of cases was enough to anchor expectations. Are we able to rule out competing hypotheses/mechanisms?

One competing hypothesis is that our results have nothing to do with the small share of cases anchoring expectations. Instead, it is simply picking up the direct effect of the Fire Court rulings. This is because we are not able to distinguish in the data properties that went to the Fire Court and those that did not. The regression therefore includes all properties in the parish – those that went to the Fire Court and those that did not. To see why this might be a problem, consider the following example of a parish where there are 100 properties (see Table 1.13). Before the Fire, the average number of hearths per property was 10. Now assume that of the 100 properties, 60 did not go to Court but 40 went to Court. Let us further assume that the average number of hearths in the 60 properties was the same before and after the Fire (i.e., 10 hearths). However, in the 40 cases that went to the court, the average number of hearths increased by three per property to 13 hearths. Consequently, the overall average number of hearths per property increased by 1.2 to 11.2. This stylized example shows us how the average number of hearths can increase even without any anchoring of expectations of those that did not go to Court.

Table 1.13: Stylized example

	Total Hearths	Number of properties	Avg. hearths per property
Before Fire	100*10=1,000	100	10
After Fire	(60*10)+(40*13)=1,120	100	11.2

However, we think that this should not affect our results substantially since the proportion of properties in each parish that went to the Fire Court is a relatively small number. Based on the initial cases, the average percentage of properties in each parish that went to the Fire Court was 6%, the median was 4% and the maximum was 30%. In addition, Table 1.14 shows the results when we drop all parishes where more than 14.6% (column 2), 7.4% (column 3), 4.1% (column 4) and 2.3% (column 5) of the properties went to the Fire Court. 14.6%, 7.4%, 4.1% and 2.3% correspond to the 95th, 75th, 50th and 25th percentiles respectively.

Column 1 shows the results using the full sample. Our results remain robust to dropping parishes where a “large” proportion of properties went to the Fire Court. If anything, our results seem to get bigger when we drop more parishes which is against what we should see if our results are purely picking up the direct effect of the Fire Court rulings.

Table 1.14: Effect of legal rulings on the number of hearths per property (dropping parishes where a "large" proportion of properties went to Court)

	(1)	(2)	(3)	(4)	(5)
VARIABLES			ln(No. Hearths per Property)		
Pragmatic X Post Fire	0.981***	0.893***	0.897***	1.051***	1.121*
	(0.246)	(0.241)	(0.245)	(0.232)	(0.527)
Observations	31,582	30,993	29,210	26,289	21,953
R-squared	0.031	0.032	0.033	0.035	0.042
Parish FE	✓	✓	✓	✓	✓
Post FE	✓	✓	✓	✓	✓
Parish controls X Post FE	✓	✓	✓	✓	✓
Broader location X Post FE	✓	✓	✓	✓	✓
Pre-fire hearth tercile X Post FE	✓	✓	✓	✓	✓
Number of clusters	46	43	34	24	12
Sample	All	< 95 pct	< 75 pct	< 50 pct	< 25 pct

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Another competing hypothesis is that of the share of owner-occupied properties in the parish. In parishes where there is a large share of owner-occupied properties, the share of properties that go to Court must by definition be small since owners cannot sue themselves. Therefore, our results may have nothing to do with the small share of cases anchoring expectations. Instead, they could simply be reflecting the fact that these parishes have a greater share of owner-occupied properties. Landowners are less likely to be credit constrained and can thus allocate more resources towards building more hearths per property. In addition, if the owner occupies the property, then there is a clear assignment/alignment of property-rights. Consequently, the owner builds at a higher quality because the owner is able to accrue the full benefits of living in a higher quality property.

Unfortunately, the data do not tell us whether a property is owner occupied – they only tell us who the main occupant of the property is. To overcome this limitation, we count the share of peers, high-ranking military personnel and doctors living in the parish. To the extent that this group of individuals are more likely to own their own homes, this variable gives us a proxy for the share of owner-occupied properties in the parish. We then include these three variables as controls in our regression. Referring to Table 1.10 column 2, we can see that the coefficient estimate on $PragmaticRulings_j \times PostFire_t$ remains stable to the inclusion of these variables as controls. This suggests that there are aspects of the rulings in the small share of properties that went to Court that cannot be attributed to the share of owner-occupied properties in the parish. Therefore, our results lend credence to our proposed mechanism that the small share of cases was enough to anchor expectations for everyone in the parish.

1.5.7 Robustness checks

Dropping parishes which merged after the Fire. One concern could be that the Fire led to the merging of some parishes and so our results could be driven by these enlarged parishes which might have more resources. In Table A.16 we re-run regression 1.11 using only parishes that did not merge after the Fire. Reassuringly, the coefficient estimates remain similar and even larger than our baseline results, suggesting that our baseline results are conservative.

Accounting for zeros in the outcome variable. For burned parishes with Fire Court cases, there are 801 observations which are recorded as having zero hearths in the property. Taking logs results in these observations dropping out of the regression. Therefore, to account for the zeros in the outcome variables, we adopt two approaches. First, applying the inverse hyperbolic sine transform to hearths. Second, using a Poisson pseudo-likelihood (PPML) regression. The results are reported in Tables A.17 and A.18 respectively. The estimated effects are very similar to our baseline results.

Trimming extreme values of the outcome variable. Another robustness check that we run is to drop the top and bottom 1 percentile of $ln\left(Hearths_{ijt}\right)$. The results of this robustness check is reported in Table A.19 and are similar to our baseline results.

1.5.8 Using an IV estimation strategy

As there could be concerns that there are time-varying parish-level omitted variables which we have not controlled for, we augment our DiD strategy with an instrumental variable (IV) strategy. Our IV strategy exploits the fact that Fire Court judging panels with different political alignments (i.e., whether they were predominantly Royalists or Parliamentarians) were assigned to the cases in the parishes. The 1666 Great Fire took place in the midst of the Second Dutch War (1665-1667) and the Great Plague which began in 1665. King Charles II was relying on taxes and loans from London and its wealthiest citizens to finance the war. The destruction of the customs house, wharves and more than 13,000 buildings caused a significant drop in royal revenue and thus the King had a vested interest for London to be quickly rebuilt. Therefore, judging panels that were predominantly Royalists (more aligned with the King) were more likely to decree pragmatic rulings so as to facilitate the rebuilding of London. As a result, we can use the composition of the judging panels as an instrument for the share of initial cases in the parish that had pragmatic rulings. This gives us exogenous variations in legal rulings for each of the parishes. Of the 46 parishes in our regression with Fire Court cases, 17 of them (37.0%) had the majority of their initial cases presided by judging panels that consisted predominantly of Royalists.

Relevance of instrument

We estimate the first-stage relationship between the composition of the judging panels in the initial cases and the share of initial cases in the parish that had pragmatic rulings:

$$\begin{aligned} PragmaticRulings_j \times PostFire_t = \alpha_j + \delta PostFire_t + \beta MajorityRoyalist_j \times PostFire_t \\ + \gamma' X_{jt} + u_{ijt} \end{aligned} \tag{1.12}$$

Table 1.15 presents the first-stage results which suggest that if the majority of the initial cases in the parish were heard by judging panels that were predominantly Royalists, then the share of initial cases in the parish that had pragmatic rulings increased by 9.5%-pts.

Table 1.15: First-stage – Effect of Royalist on legal rulings

	(1)	(2)	(3)	(4)
VARIABLES		Pragmatic X Post		
Majority royalist in judging panels X Post	0.119***	0.087*	0.095***	0.095***
	(0.038)	(0.045)	(0.031)	(0.029)
Observations	31,582	31,582	31,582	31,582
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46
KP F-stat	9.895	3.795	9.122	10.37

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

The first-stage relationship is robust to the inclusion of a series of parish controls interacted with $PostFire_t$, broader locations-by-post fixed effects and pre-Fire hearth terciles-by-post fixed effects. In addition, in most of the regressions, the first-stage has a KP F-statistic value of around 10. Figure A.4 shows the binned scatter plot of the results in column 4.

Validity of instrument

Conditional Independence. The validity of our instrument depends crucially on whether there were other parish-level factors involved in determining the composition of the judging panels in the initial cases. We verify this by running a balancing test. This is similar to the type of statistical test that is used to verify random assignment in a randomized controlled trial.

Table 1.16: Testing for random assignment of judging panels to parishes

	(1)	(2)	(3)	(4)
VARIABLES	Majority royalist in judging panels			
ln(No. properties before the Fire)	-0.033	0.004	0.034	0.029
	(0.064)	(0.086)	(0.095)	(0.133)
Share of peers	-2.071	-3.816	-6.501	-4.256
	(9.929)	(10.247)	(11.106)	(23.191)
Share of high-ranking military personnel	-30.800	-7.468	-10.069	-21.936
	(29.296)	(42.349)	(51.972)	(81.006)
Share of doctors	1.689	7.120	2.283	-7.803
	(17.258)	(19.069)	(19.263)	(27.936)
Broader location 1		-0.097	-0.075	0.098
		(0.324)	(0.322)	(0.494)
Broader location 2		0.161	0.169	0.220
		(0.288)	(0.300)	(0.518)
Pre-Fire hearth tercile 1			-0.177	-0.315
			(0.236)	(0.292)
Pre-Fire hearth tercile 2			-0.049	-0.063
			(0.210)	(0.250)
ln(Average rent in 1638)				-0.175
				(0.289)
Observations	46	46	46	37
Adjusted R-squared	-0.081	-0.109	-0.152	-0.259
F-stat for joint test	0.350	0.431	0.373	0.361
p-value for joint test	0.843	0.853	0.928	0.932

Notes: Robust standard errors. Notation for statistical significance:
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.16 shows the result of this balancing test. The coefficients on the share of peers, high-ranking military personnel and doctors seem sizable. However, this is because the mean values of these variables are extremely small. For example, the mean value for the share of peers is 0.005, that of high-ranking military personnel is 0.0004 and that for doctors is 0.003.

Table 1.16 also shows that parish-level factors were not predictive of the composition of the judging panels in the initial cases. In column 4, for parishes where data were available, we included the average rent in the parish in 1638. The 1638 rental data comes from "The Inhabitants of London in 1638".[16] Column 4 shows that historical rents were not predictive of the composition of the judging panels. Importantly, across all of the columns, all of the estimates are statistically insignificant at the 1% level and are not jointly significant with p-values ranging from 0.84 to 0.93.

Exclusion. This restriction requires that the composition of the judging panels in the initial cases affected the quality of building investment in the parish only through its effect on legal rulings. While it is not possible to formally test the exclusion condition, the fact that our instrument passes the balancing test is reassuring.

However, there could still be concerns that the exclusion restriction could be violated since it is possible that monarchist officials may had some influence in the assignment of judges to the cases. For example, these officials could be expecting some parishes to grow more, and so they wanted to make sure that the parliamentarian judges did not derail their plans. In addition, it could be the case that some parishes had more monarchist landowners and so the officials assigned monarchist judges to protect the interest of these landowners.

In order to address such concerns, we are in the process of collecting data to measure what is the share of peers in each parish that was loyal to the King. From the hearth tax data, we are able to identify the names of the peers (i.e., Duke, Duchess, Marquess, Marchioness, Earl, Countess, Viscount, Viscountess, Baron, Baroness, Lord, Lady, Sir, Dame and Ambassador) living in each parish. For the parishes in our regression, there are about 500 peers in total. We can then refer

[16]"The Inhabitants of London in 1638" was originally published by the Society of Genealogists in London in 1931 (see Dale (1931)) and can be accessed at the British History Online website. This publication was based on the manuscript "Settlement of Tithes, 1638", found in the Lambeth Palace Library. The manuscript contains a list of the householders in 93 out of the 107 parishes in the City of London, as well as the rentals paid for the houses and the tithes paid.

to various historical sources (e.g., the Oxford Dictionary of National Biography (2004)) to find out what were the views of these peers on the 1660 restoration of the monarchy. Once we have determined the share of monarchist peers in each parish, we can then include this variable as a control in the regression. This would hopefully help to control for the possibility that monarchist judges were being assigned to parishes where the monarchists wanted to have greater influence over or benefit more from.

In any case, even if the exclusion restriction is violated, our reduced form estimates can still be interpreted as the causal effect of the composition of the judging panels in the initial cases on the number of hearths per property in the parish.

Monotonicity. The monotonicity assumption requires a monotonic relationship between the instrument and the variable that is being instrumented. The monotonicity assumption ensures that our IV estimate can be interpreted as a local average treatment effect (LATE). In our context, this is the average causal effect among the subgroup of parishes that invested differently in their buildings because of the composition of the judging panels in the initial cases.

If the monotonicity assumption is violated, then in the classical IV framework, our results can only be interpreted as causal **constant** effects. However, in a heterogeneous treatment effects framework, if the monotonicity assumption is violated, Angrist et al. (1996) and Heckman and Vytlacil (2005) show that the IV estimates would still be a weighted average of marginal treatment effects. However, because the weights do not sum to one, this leads to an ill-defined local average treatment effect.

One testable implication of the monotonicity assumption is that the first-stage estimates should be non-negative for any subsample. To test this, we split the sample into various subsamples and estimated the first-stage relationship for each of these subsamples. The results are reported in Table A.20. In columns 1 and 2, we split the sample into whether the church in the parish was damaged by the Fire. In columns 3 to 5, we split the sample into three broader geographical locations (i.e., abutting the City of London walls, within the walls and outside the walls). In columns 6 to 8, we split the sample based on terciles of the number of hearths in each parish before the Fire. In columns 9 to 11, we split the sample based on terciles of the number of properties in each parish before the Fire. In columns 12 to 14, we split the sample based on terciles of the share of peers in the parishes. In columns 15 and 16, we split the sample into two based on the share of doctors in the parishes. Finally, in columns 17 and 18, we split the sample into two based on the share of high-ranking military personnel in the parishes. Out of these 18 subsamples, there are only three subsamples where the first-stage estimate is negative. In the other 15 subsamples, the first-stage estimates are positive, consistent with the monotonicity assumption.

IV results and discussion

Table 1.17 reports the results from the IV regressions. In column 4, the results suggest that parishes where all the initial cases saw pragmatic Fire Court rulings experienced a highly statistically significant increase of around 200% more hearths compared to parishes where all the initial cases saw unpragmatic rulings. In other words, going from the 25th percentile parish in terms of the share of pragmatic rulings (0% pragmatic rulings) to the 75th percentile parish (20% pragmatic rulings) resulted in a 40% increase in the number of hearths per property. In absolute terms, this corresponds to an increase of 1.53 hearths. We also report the reduced form estimates in Table A.21 and show the associated binned scatter plot of the residues (based on all the controls) in Figure A.5.

The IV results are around twice as large as the DiD results (1.53 hearths vs. 0.75 hearths). This could suggest two things. First, the DiD regression suffers from omitted variables and thus fails the parallel trend assumption. Second, even if the DiD estimate is unbiased, we should still expect the IV estimate to be different from it. This is because the DiD estimate identifies the average treatment effect whereas the IV estimate gives us the local average treatment effect for the compilers. Nevertheless, the fact that the IV estimates are positive and highly statistically significant re-affirms our DiD results. This gives us greater confidence in concluding that pragmatic legal rulings affected individuals' expectations about how much other individuals in their parish would be investing. This in turn resulted in a higher number of hearths per property in the parish.

Table 1.17: IV – Effect of legal rulings on the number of hearths per property

	(1)	(2)	(3)	(4)
VARIABLES	ln(No. hearths)			
Pragmatic X Post Fire	2.276***	2.550***	2.247***	2.001**
	(0.698)	(0.917)	(0.753)	(0.789)
Observations	31,582	31,582	31,582	31,582
R-squared	0.010	0.016	0.020	0.026
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46
KP F-stat	9.895	3.795	9.122	10.37

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Adding in other dimensions of the rulings as controls. To address concerns that our definition of pragmatic rulings fails to consider changes in rent and tenancy length, we include these as controls in our regression. Table A.22 provides evidence that it is indeed pragmatic Fire Court rulings that are affecting individuals' expectations and not the other dimensions of the Fire Court's decisions.

Robustness checks

Dropping parishes which merged after the Fire. Similar to the other sections, in Table A.23 we re-run our IV analysis using only parishes that did not merge after the Fire. The coefficient estimates remain very similar to our baseline results.

Accounting for zeros in the outcome variable. To account for the zeros in the outcome variables, we apply the inverse hyperbolic sine transform to hearths. The results are reported in Table A.24. The estimated effects remain positive and continue to be highly statistically significant.

Trimming extreme values of the outcome variable. Table A.25 reports the results where we drop the top and bottom 1 percentile of $ln\left(Hearths_{ijt}\right)$. The results are similar to our baseline results.

1.6 Conclusion

The development of cities often involves the rejuvenation and replacement of outdated buildings. However, the sunk cost of existing durable structures often serves as an impediment. While disasters are destructive, an unintended silver lining is that they may help to remove development frictions. By lowering the opportunity cost of waiting to rebuild to zero, disasters could potentially spur the development of neighborhoods and even cities. However, disasters do not necessarily guarantee higher quality buildings. What ultimately matters is what each individual expects other individuals to do. Our paper highlights this by providing causal evidence of how legal rulings can be a main driver in the formation of these expectations. While there is a relatively large theoretical literature on how legal institutions can affect expectations and hence the behavior of individuals, there is relatively less empirical work on this. Our paper thus addresses this gap in the literature. Although the setting of our paper is 17th century England, even today, legal rulings continue to be a key aspect in society. This has policy implications as it suggests scope for laws to influence expectations and in so doing, facilitate the continual development of cities.

Chapter 2: Made by History: The Spatial Distribution of Manufacturing Industries in the US

2.1 Introduction

There is an extensive literature that shows that the quirks of history shape present-day economic outcomes either through path dependency and/or through altering long-run determinants of these outcomes. One famous example is Bleakley and Lin (2012) where they show that even though portage is now an obsolete technology, there is still greater population density today in areas near portage sites. Recently, a ground-breaking paper by Bazzi et al. (2020) contributes to this literature by showing how time spent on the American frontier has contemporaneous and persistent effects on a location's culture and institutions. Specifically, "rugged individualism" continues to persist in counties that spent more time on the frontier. Residents of these counties exhibit more pervasive individualism, prefer less redistribution, lower public spending and less social protection in terms of minimum wages, gun control and environment protection.

In this paper, I build on this existing literature by examining the following **research question** – how does this particular episode of history (time at the frontier) help to explain manufacturing production patterns across American counties today? In other words, can time at the frontier explain the present-day sorting pattern of specific industries to specific counties? Following Nunn (2007b) and Alfaro et al. (2019), I define industries based on how susceptible they are to holdup (i.e., level of "contractibility").

I begin this paper with a simple conceptual framework to explain how time spent on the frontier affects the spatial distribution of manufacturing industries across US counties. The model is adapted from Melitz (2003). Instead of selection into exporting, I adapt the model to show how firms and industries sort into producing at different locations. In the model, there is a continuum of firms with varying levels of "contractibility". Higher "contractibility" means that the intermediate inputs used by the firm are less relationship-specific and hence easier to specify in contracts. In other words, these industries are less susceptible to holdup. This definition is based on Nunn (2007b) and Alfaro et al. (2019). Given their level of "contractibility", firms decide which counties to produce in. There is a fixed cost associated with producing in each county. In particular, the more time that a county spent on the frontier, the higher is the fixed cost that the firm needs to incur to produce there. There are various interpretations for this higher cost. One could be that the time spent on the frontier is associated with the culture and institutions of the county being more individualistic and these translate to a higher cost. Such an interpretation is consistent with Bazzi et al. (2020).

This simple framework yields the following implications. First, the outcomes at the county level. There are fewer firms in counties that spent a longer time on the frontier. However, this is ambiguous for employment. The intuition is that producing in counties that spent a longer time on the frontier is less profitable due to the high fixed cost associated with the location. Second, the outcomes at the industry level. Under certain distributional assumptions, there are fewer firms in industries that are more "contractible". The intuition is that there is a lower probability of a firm drawing a high "contractibility" parameter. Third, the sorting pattern of industries to counties. Firms in high "contractibility" industries sort into producing at counties that spent a longer time on the frontier. However, this is ambiguous for employment. The intuition for this is that "contractibility" is somewhat akin to productivity – only the high "contractibility" firms are able to pay the high fixed cost associated with producing in counties that spent a longer time on the frontier.

I begin my empirical analysis by first showing evidence of the first two implications of the model. I show that there are indeed fewer establishments and lower employment in counties that spent a longer time on the frontier. This result is robust to the inclusion of a rich set of geographical and historical controls. In addition, I show that there are fewer establishments and lower employment in industries that are more "contractible". However, the estimates are imprecisely estimated.

Next, I examine the present-day sorting pattern of specific industries to specific counties. This is the main contribution of my paper. To do this, I employ a difference-in-differences (DiD) strategy which exploits two cross-sectional variations. The first cross-sectional variation comes from the fact that different counties spent different number of years on the American frontier. The second cross-sectional variation arises because different industries have different levels of "contractibility". For example, industries that predominantly use inputs that are sold on an organized exchange (e.g., oil) are highly "contractible". This means that the inputs are less relationship-specific and hence easier to specify in contracts. Using this DiD strategy, I find that the time spent on the frontier does indeed affect the present-day composition of manufacturing industries across US counties. Counties that spent a longer time on the frontier have relatively more establishments and higher employment in "contractible" industries today. The results are consistent with the mechanisms proposed in the conceptual framework – only the high "contractibility" firms are able to pay the high fixed cost associated with producing in counties that spent a longer time on the frontier. The results hold even when a rich set of geographical, historical and industry controls, as well as extremely demanding fixed effects are included. This rules out the role of geography, other historical events and other industry characteristics as alternative explanations and hence confounders to the results. In addition, the results continue to hold even when using the Herfindahl index of intermediate input use as an alternative measure of holdup in each industry.

Having shown strong evidence of the sorting pattern of high "contractibility" industries to counties that spent a longer time on the frontier, I next explore various mechanisms that could explain this sorting pattern. The first mechanism that I explore is that of the first-mover advantage that persists due to path dependency. Under this explanation, the present-day high "contractibility" industries that sort to counties that spent a longer time on the frontier are the same industries that sort to these counties in the past. Due to path dependency, this sorting pattern continues to persist till today even if the initial reasons for why these industries locate in these counties are no longer relevant. I show evidence that while the high "contractibility" industries of the past did sort to counties that spent a longer time on the frontier, these industries are different from the present-day high "contractibility" industries. Since the set of industries that sort to the frontier counties are different in the two time periods, this allows me to rule out the first-mover advantage mechanism.

The second mechanism that I explore is whether the present-day sorting pattern can be explained by the culture and institutions of frontier counties. The idea behind this mechanism stems from the finding in Bazzi et al. (2020) that the time spent on the frontier has both contemporaneous and persistent effects on the counties' culture and institutions. This suggests that the individualistic culture and institutions that developed in the early days became fundamental to a place over time and hence persist till today. Relying on the arguments in Bazzi et al. (2020), I argue that the "rugged individualism" of a county affects the sorting of industries. Since counties that spent a longer time on the frontier have more individualistic culture and institutions, individuals in these counties are less likely to trust other people. Therefore, anything that is not "contractible" becomes harder and more costly to enforce. Consequently, only the more "contractible" industries locate in counties that spent a longer time on the frontier.

This paper is related to the burgeoning literature on the persistence of history. Many of these papers are related to the literature on the effect of institutions on economic outcomes which I briefly cover in the paragraphs to come. Besides those papers, other examples include Nunn (2007a, 2008), Bleakley and Lin (2012), Hornbeck and Naidu (2014), Jedwab and Moradi (2016), Sequeira et al. (2020), Allen and Donaldson (2020), Bazzi et al. (2020) and Ottinger (2020). Related to the literature on the persistence of history are papers that examine whether major shocks are able to break this persistence. These shocks include wars and bombing (Davis and Weinstein (2002) and Miguel and Roland (2011)), natural or man-made disasters (Siodla (2015) and Hornbeck and Keniston (2017)), political events (Redding et al. (2011) and Michaels and Rauch (2018)), technology (Bleakley and Lin (2012)) and even diseases (Jedwab et al. (2019)). This paper builds on Bazzi et al. (2020) and shows how time spent at the frontier has persistence consequences on how industries sort to counties.

In addition, this paper is related to the literature on what explains the spatial distribution of economic activities. Economic geography theories suggest that this is determined by what Fujita and Thisse (2002) describe as the "fundamental trade-off of urban economics" involving agglomeration versus dispersion forces. The existing literature has mainly studied these forces in terms of the externalities described in Marshall (1890) (i.e., labor-market interactions, linkages between firms and knowledge spillovers),[1] transportation infrastructure, endogenous amenities and government policies.[2] There have also been papers that examine the role of location fundamentals (i.e., geography and factor endowments) in explaining the spatial distribution of economic activities. These papers include Fuchs (1962), Kim (1995), Ellison and Glaeser (1999), Hanson and Slaughter (2002) and Davis and Weinstein (2002).

In examining the sorting patterns of industries to counties, this paper is related to the literature on the sources of comparative advantage. Nunn and Trefler (2014) document that there has been a growing literature that explores whether institutions are a source of comparative advantage. These include contracting institutions in the product market (Nunn (2007b), Levchenko (2007), Boehm and Oberfield (2020) and Boehm (2020)), labor market related institutions (Costinot (2009) and Cunat and Melitz (2012)) and financial institutions (Beck (2003) and Manova (2008, 2013)). A recent paper by Chor (2010) simultaneously tests all of these different institutions by including all

[1]These Marshallian externalities are similar to what Duranton and Puga (2004) describe as sharing, matching and learning.

[2]Examples of papers examining Marshallian externalities are Holmes (1999), Costa and Kahn (2000), Henderson (2003), Moretti (2004), Ellison et al. (2010), Greenstone et al. (2010), Dauth et al. (2018) and Atkin et al. (2019). Examples of papers examining transportation infrastructure are Baum-Snow (2007), Michaels (2008), Duranton and Puga (2012), Faber (2014), Donaldson and Hornbeck (2016), Donaldson (2018) and Heblich et al. (2020). As for endogenous amenities, examples are Diamond (2016), Davis et al. (2019) and Couture et al. (2020). An example of a paper studying the role of government policies is Kline and Moretti (2013).

of them in a single regression. Finally, there has been a recent literature that looks specifically at the comparative advantage of more granular geography such as cities. Examples include Gaubert (2018), Tian (2019) and Davis and Dingel (2020). This paper contributes to this literature by showing how high "contractibility" industries sort to counties that spent a longer time on the frontier.

By arguing that the time spent on the frontier affects a county's culture and institutions, and this consequently affects the sorting of industries to counties, this paper is related to the literature on the effect of institutions on economic outcomes. North and Thomas (1973) argue that the commonly studied determinants of growth such as innovation, economies of scale, education and capital accumulation "are not causes of growth; they are growth". In their view, the "fundamental" determinant of growth is institutions. This has led to a burgeoning literature which include Engerman and Sokoloff (1997), La Porta et al. (1997, 1998), Acemoglu et al. (2001, 2005), Banerjee and Iyer (2005), Dell (2010), Michalopoulos and Pappaioannou (2016), Dell and Olken (2020) and Mendez-Chacon and Patten (2020).[3]

While this paper shows that history clearly matters, a lingering question remains. "Does history matters only when it matters little" (Rauch (1993))? A ground-breaking paper by Allen and Donaldson (2020) provides compelling evidence that this seems to be the case. Their paper shows that while history determines where economic activities take place, it plays a very limited role in determining overall welfare in the long run. However, if we consider that history also affects the "fundamental" determinants of welfare such as culture and institutions (a channel which is not modelled in Allen and Donaldson (2020)), then it might be the case that history actually plays a bigger role in affecting welfare. Examining this is left to future work.

[3]There is also a distinct literature that examines the effect of culture on economic outcomes. Examples of these include Weber (1930), Fischer (1989), Greif (1994), Zerbe and Anderson (2001), Guiso et al. (2016), Tabellini (2010), Nunn and Wantchekon (2011), Alesina et al. (2013) and Gorodnichenko and Roland (2017).

The rest of the paper proceeds as follows. Section 2.2 presents a simple conceptual framework which shows how time spent on the frontier affects the spatial distribution of economic activities. Section 2.3 presents the historical background as well as the data used in the main analysis. Section 2.4 examines the first two implications of the model – the number of establishments and employment at the county level, as well as at the industry level. Section 2.5 examines the sorting pattern of industries to counties. In particular, whether high "contractibility" industries sort to counties that spent a longer time on the frontier. Section 2.6 explores potential explanations behind this sorting pattern. Finally, I conclude in Section 2.7.

2.2 Conceptual framework

2.2.1 Overview

In this section, I present a simple conceptual framework to explain how time spent on the frontier affects the spatial distribution of manufacturing industries. This model is adapted from Melitz (2003). Instead of selection into exporting, I adapt the model to show how firms and industries sort into producing at different locations. In the model, there is a continuum of firms with varying levels of "contractibility" (i.e., susceptibility to holdup). Higher "contractibility" means that the intermediate inputs used by the firm are less relationship-specific and hence easier to specify in contracts. This definition is based on Nunn (2007b) and Alfaro et al. (2019). Given their level of "contractibility", firms decide which counties to produce in. There is a fixed cost associated with producing in each county. The fixed cost varies with the amount of time that the county spent on the frontier. In particular, the longer the county spent on the frontier, the higher is the fixed cost that the firm needs to incur to produce there. There are various interpretations for this higher cost. One could be that the time spent on the frontier is associated with the culture and institutions of the county being more individualistic and these translate to a higher cost when it comes to negotiating with suppliers. In Appendix B.1.1, I show how this can be microfounded using a model of sequential production.[4] In addition, I further examine this interpretation in Section 6 where I explore potential mechanisms.

2.2.2 Consumers

Consumers are homogeneous and have CES preferences over varieties. A representative consumer in county c gets utility U_c from the consumption of goods produced in his/her own county, as well as goods shipped by other firms in other counties. Ω denotes the continuum of possible goods that can be produced. Each good $\omega \in \Omega$ is defined by the firm and its location. For example, a firm that produces at two different locations is considered to be producing two different goods. Let the income of county c be denoted by Y_c, the price of a good from county r consumed in county c be p_{rc} and the quantity of a good from county r consumed in county c be q_{rc}. The representative consumer's problem in each county is:

$$\underset{\{q_{rc}(\omega)\}_{\omega \in \Omega}}{max} \; U_c = \left[\sum_{r=1}^{\bar{C}} \int_{\Omega_r} q_{rc}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}} \quad , \;\; \sigma > 1$$

subject to the budget constraint $\sum_{r=1}^{\bar{C}} \int_{\Omega_r} p_{rc}(\omega) q_{rc}(\omega) d\omega \leq Y_c$.

Solving this maximization problem yields the optimum quantity demanded of good ω:

$$q_{rc}(\omega) = p_{rc}(\omega)^{-\sigma} Y_c P_c^{\sigma-1}$$

where $P_c \equiv \left[\sum_{r=1}^{\bar{C}} \int_{\Omega_r} p_{rc}(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}$ is the Dixit-Stiglitz price index.

[4]In this microfoundation, the production of final goods entails a continuum of intermediate stages. In each stage, final good producers bargain with suppliers of intermediate inputs. However, due to incomplete contracts, holdup occurs. In counties with more individualistic culture and institutions, this holdup is more severe, leading to a higher fixed cost of production.

2.2.3 Firms

The timing of decisions for the firm is as follows. First, the firm decides whether to pay a fixed entry cost (f_e) which entitles it to obtain a "contractibility" draw (ψ). I denote the probability density function that ψ is drawn from as $g(\psi)$ and the cumulative distribution function as $G(\psi)$. Second, the value of ψ that was drawn determines the industry of the firm. Third, having observed ψ, the firm decides whether or not to produce in each of the counties. Finally, the firm decides on the quantity to produce. The problem is solved by backward induction. Going forward, I index firms by ψ.

Deciding on the quantity to produce

There are $\bar{C}$ counties with varying levels of time spent on the frontier (ϕ_c). Counties are indexed 1 to $\bar{C}$ with, 1 reflecting the county that spent the shortest amount of time on the frontier and $\bar{C}$ denoting the county that spent the longest amount of time on the frontier. Let $\tau_{rc} \geq 1$ denote the iceberg trade cost. The problem of a firm in county c is:

$$\underset{\{y_{rc}(\psi)\}_{\psi \in \Psi}}{max} \sum_{r=1}^{\bar{C}} \left(p_{rc}(\psi) y_{rc}(\psi) - \frac{w_c \tau_{rc}}{\psi} y_c(\psi) \right) - f(\phi_c)$$

subject to the demand for each good $q_{rc}(\omega) = p_{rc}(\omega)^{-\sigma} Y_c P_c^{\sigma-1}$ and the goods market clearing condition $y_{rc}(\psi) = q_{rc}(\psi)$. w_c is the wage in county c and $f(\phi_c)$ is the fixed cost associated with producing in county c. I assume that the longer the county spent on the frontier, the higher is the fixed cost that the firm needs to incur to produce there. Therefore, $\frac{\partial f(\phi_c)}{\partial \phi_c} > 0$. The amount of labor used to produce $y_{rc}(\psi)$ is $\frac{\tau_{rc} y_{rc}(\psi)}{\psi} + f(\phi_c)$.

Solving this maximization problem yields:

$$p_{rc}(\psi) = \frac{\sigma}{\sigma - 1} \frac{w_c \tau_{rc}}{\psi}$$

The profits of the firm in county c is thus:

$$\pi_c(\psi) = B_c \psi^{\sigma-1} - f(\phi_c)$$

where $B_c \equiv \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\sum_{r=1}^{\bar{C}}\left(w_c\tau_{rc}\right)^{1-\sigma} Y_r P_r^{\sigma-1}$.

Deciding whether or not to produce in each of the counties

The firm will produce in county c as long as:

$$\pi_c(\psi) \geq 0$$

Let $\underline{\psi}_c$ be the value of ψ where $\pi_c(\psi) \geq 0$. This gives us the **zero cutoff "contractibility" condition** for each county:

$$\psi \geq \underline{\psi}_c \equiv \frac{f(\phi_c)\sigma^{\sigma}(\sigma-1)^{1-\sigma}}{\sum_{r=1}^{\bar{C}}\left(w_c\tau_{rc}\right)^{1-\sigma} Y_r P_r^{\sigma-1}}$$

Note that the cutoff is increasing in the amount of time spent on the frontier since:

$$\frac{\partial \underline{\psi}_c}{\partial \phi_c} = \frac{\sigma^{\sigma}(\sigma-1)^{1-\sigma}}{\sum_{r=1}^{\bar{C}}\left(w_c\tau_{rc}\right)^{1-\sigma} Y_r P_r^{\sigma-1}}\frac{\partial f(\phi_c)}{\partial \phi_c} > 0$$

In addition, any firm that draws $\psi < \underline{\psi}_1$ will not produce at all.

Determining the industry of the firm

There are J industries with varying levels of "contractibility" $(\tilde{\psi}_j)$. Industries are indexed 1 to J with, 1 reflecting the industry which is the least "contractible" and J denoting the industry which is the most "contractible". To recap, higher "contractibility" means that the intermediate inputs used by the firm are less relationship-specific and hence easier to specify in contracts.

The firm's industry is determined exogenously based on the "contractibility" value (ψ) that it drew. In particular, a firm will be in industry j if:

$$\tilde{\psi}_j \leq \psi < \tilde{\psi}_{j+1}$$

Deciding whether to pay a fixed entry cost to obtain a "contractibility" draw

In order to obtain a "contractibility" draw, the firm needs to pay a fixed entry cost f_e. This entry cost is paid in terms of labor and once paid, allows the firm to enter in all locations. The firm will choose to obtain a "contractibility" draw if the expected profits cover the entry cost:

$$\begin{aligned}\int_0^\infty \sum_{s=1}^{\bar{C}} \pi_s(\psi) g(\psi) d\psi = &\int_{\underline{\psi}_1}^\infty \sum_{s=1}^{\bar{C}} (B_s \psi^{\sigma-1} - f(\phi_s)) g(\psi) d\psi \\ =&f_e\end{aligned}$$

This gives us the **free-entry condition**.

The expected profits conditional on entering (i.e., obtaining a "contractibility" draw) is:

$$\begin{aligned}\widetilde{\pi}(\psi) =& \int_{\underline{\psi}_1}^\infty \sum_{s=1}^{\bar{C}} \pi_s(\psi) \frac{g(\psi)}{1 - G(\underline{\psi}_1)} d\psi \\ =&\frac{1}{1 - G(\underline{\psi}_1)} \int_{\underline{\psi}_c}^\infty \sum_{s=1}^{\bar{C}} \pi_s(\psi) g(\psi) d\psi \\ =&\frac{f_e}{1 - G(\underline{\psi}_1)}\end{aligned}$$

To simplify the math, I assume $\tau_{rc} = \tau_c \geq 1$. Consequently, the number of firms producing in the entire country is:

$$M = \frac{L}{\left[\frac{\sigma f_e}{1 - G(\underline{\psi}_1)} + \sum_{s=1}^{\bar{C}} \frac{\sigma + w_s - 1}{w_s} f(\phi_s)\right]} \tag{2.1}$$

The full derivation of this equation can be found in Appendix B.1.2. Note that M is decreasing in the level of individualistic institutions:

$$\frac{\partial M}{\partial \phi_s} = -\,L\left[\frac{\sigma f_e}{1-G(\underline{\psi}_1)} + \sum_{s=1}^{\bar{C}} \frac{\sigma + w_s - 1}{w_s} f(\phi_s)\right]^{-2}$$
$$\left[\frac{\sigma + w_s - 1}{w_s}\frac{\partial f(\phi_s)}{\partial \phi_s} + \mathbf{1}(s=1)\sigma f_e \left[1-G(\underline{\psi}_1)\right]^{-2} \frac{\partial G(\underline{\psi}_1)}{\partial \phi_1}\right] < 0$$

since $\frac{\partial f(\phi_s)}{\partial \phi_s} > 0$ and $\frac{\partial G(\underline{\psi}_s)}{\partial \phi_s} = \frac{\partial G(\underline{\psi}_s)}{\partial \underline{\psi}_s}\frac{\partial \underline{\psi}_s}{\partial f(\phi_s)}\frac{\partial f(\phi_s)}{\partial \phi_s} > 0$.

Now, consider the case where the "contractibility" of any two industries j and $j+1$, lie between the zero cutoff "contractibility" of two counties c and $c+1$:

$$\underline{\psi}_c \leq \tilde{\psi}_j < \tilde{\psi}_{j+1} \leq \underline{\psi}_{c+1}$$

This condition ensures that a county produces goods from at least two different industries. In addition, the condition is fairly innocuous since we can always define industries more finely such that this condition holds.

The number of firms producing in county c is:

$$M_c = \left[1 - G(\underline{\psi}_c)\right] M$$

The number of firms producing in county c that are from industry j is thus:

$$\begin{aligned} M_{cj} &= \left[G(\tilde{\psi}_{j+1}) - G(\tilde{\psi}_j)\right] M_c \\ &= \left[G(\tilde{\psi}_{j+1}) - G(\tilde{\psi}_j)\right]\left[1 - G(\underline{\psi}_c)\right] M \end{aligned}$$

Employment used in the production of goods (and not entry) in county c is:

$$L_c = M_c J_c$$

where $J_c = \int_{\underline{\psi}_c}^{\infty} \left(\frac{\tau_c y_c(\psi)}{\psi} + f(\phi_c) \right) \frac{g(\psi)}{1-G(\underline{\psi}_c)} d\psi$.

Employment used in the production of goods (and not entry) in industry j in county c is hence:

$$\begin{aligned} L_{cj} &= \left[G(\tilde{\psi}_{j+1}) - G(\tilde{\psi}_j) \right] L_c \\ &= \left[G(\tilde{\psi}_{j+1}) - G(\tilde{\psi}_j) \right] M_c J_c \end{aligned}$$

2.2.4 Comparative statics

Outcomes at the county level: There are fewer firms in counties that spent a longer time on the frontier. However, this is ambiguous for employment.

$$\frac{\partial M_{cj}}{\partial \phi_c} = \left[G(\tilde{\psi}_{j+1}) - G(\tilde{\psi}_j) \right] \left\{ \left[1 - G(\underline{\psi}_c) \right] \frac{\partial M}{\partial \phi_c} - M \frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} \right\} < 0$$

$$\begin{aligned} \frac{\partial L_{cj}}{\partial \phi_c} &= \left[G(\tilde{\psi}_{j+1}) - G(\tilde{\psi}_j) \right] \left[M_c \frac{\partial J_c}{\partial \phi_c} + J_c \frac{\partial M_c}{\partial \phi_c} \right] \\ &= \left[G(\tilde{\psi}_{j+1}) - G(\tilde{\psi}_j) \right] \left\{ M_c \frac{\partial J_c}{\partial \phi_c} + J_c \left[1 - G(\underline{\psi}_c) \right] \frac{\partial M}{\partial \phi_c} - M \frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} \right\} \\ &\lesseqgtr 0 \end{aligned}$$

The sign for $\frac{\partial M_{cj}}{\partial \phi_c}$ is negative because $\frac{\partial M}{\partial \phi_c} < 0$ and $\frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} > 0$. The intuition for $\frac{\partial M_{cj}}{\partial \phi_c} < 0$ is that producing in counties that spent a longer time on the frontier is less profitable due to the high fixed cost associated with producing there. Therefore, less firms choose to produce in these counties.

However, the sign for $\frac{\partial L_{cj}}{\partial \phi_c}$ is ambiguous.[5] This is because when the time spent by the county on the frontier increases, two opposing effects happen. First, the number of firms decreases ($J_c \left[1 - G(\underline{\psi}_c)\right] \frac{\partial M}{\partial \phi_c} - M \frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} < 0$ and the second line of $\frac{\partial J_c}{\partial \phi_c}$ in Appendix B.1.3). Since there are fewer firms, employment is also lower. However, there is a second opposing effect. The reduced number of firms that are producing are now hiring more labor (the third line of $\frac{\partial J_c}{\partial \phi_c}$ in Appendix B.1.3). This is because with a longer time spent on the frontier, the fixed cost associated with producing in the county increases and so more labor needs to be hired to pay this increased fixed cost.

Outcomes at the industry level: Under certain distributional assumptions, there are fewer firms in industries that are more "contractible".

$$\frac{\partial M_{cj}}{\partial \tilde{\psi}_k} = M_c \left[\frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j+1} - \frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j} \right] < 0$$

$$\frac{\partial L_{cj}}{\partial \tilde{\psi}_k} = L_c \left[\frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j+1} - \frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j} \right] < 0$$

Assume that due to an exogenous shock, all industries increase in their "contractibility" by a fixed amount. As long as $\left(\frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j+1} - \frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j} \right) < 0$, then the sign of both $\frac{\partial M_{cj}}{\partial \tilde{\psi}_k}$ and $\frac{\partial L_{cj}}{\partial \tilde{\psi}_k}$ are negative. This condition is satisfied for distributions where $\frac{\partial g(\psi)}{\partial \psi} < 0$. In other words, the tail of the distribution becomes monotonically thinner for larger values of ψ. Examples of such distributions include the Pareto distribution and a truncated normal distribution. Figure B1 plots the relationship between the number of firms in each industry against the "contractibility" of industries and shows that this a fairly reasonable assumption. Therefore, $\frac{\partial M_{cj}}{\partial \tilde{\psi}_k} < 0$ and $\frac{\partial L_{cj}}{\partial \tilde{\psi}_k} < 0$ because there is a lower probability of a firm drawing a high "contractibility" parameter.

[5] The derivation of $\frac{\partial J_c}{\partial \phi_c}$ can be found in Appendix B.1.3.

The sorting of industries to counties: Firms in high "contractibility" industries sort into producing at counties that spent a longer time on the frontier. However, this is ambiguous for employment.

$$\frac{\partial M_{cj}}{\partial \tilde{\psi}_k \partial \phi_c} = \left[\frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j+1} - \frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j} \right] \left\{ \left[1 - G(\underline{\psi}_c) \right] \frac{\partial M}{\partial \phi_c} - M \frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} \right\} > 0$$

$$\begin{aligned} \frac{\partial L_{cj}}{\partial \tilde{\psi}_k \partial \phi_c} = & \left[\frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j+1} - \frac{\partial G(\tilde{\psi}_k)}{\partial \tilde{\psi}_k} \mid_{k=j} \right] \left\{ M_c \frac{\partial J_c}{\partial \phi_c} + J_c \left[1 - G(\underline{\psi}_c) \right] \frac{\partial M}{\partial \phi_c} - M \frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} \right\} \\ & + \left[G(\tilde{\psi}_{j+1}) - G(\tilde{\psi}_j) \right] \left[\frac{\partial J_c}{\partial \phi_c} \frac{\partial M_c}{\partial \phi_c} + J_c \left[1 - G(\underline{\psi}_c) \right] \frac{\partial^2 M}{\partial \phi_c^2} - \frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} \frac{\partial M}{\partial \phi_c} \right] \\ & \lesseqgtr 0 \end{aligned}$$

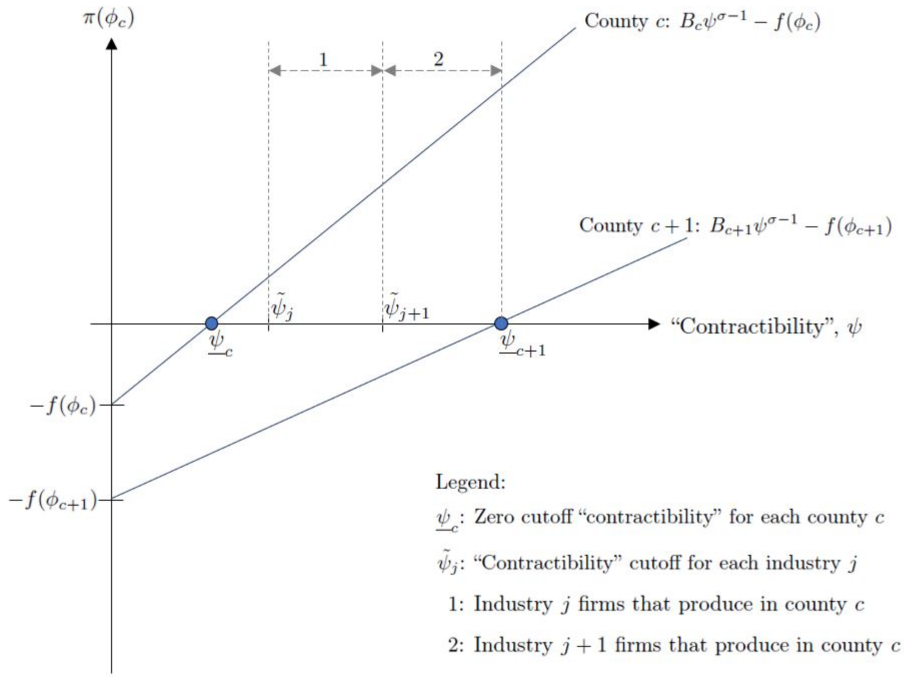
The intuition for $\frac{\partial M_{cj}}{\partial \tilde{\psi}_k \partial \phi_c} < 0$ is that "contractibility" is somewhat akin to productivity – only the high "contractibility" firms are able to pay the high fixed cost associated with producing in counties that spent a longer time on the frontier.

However, the sign for $\frac{\partial L_{cj}}{\partial \phi_c}$ is ambiguous even if we were to make certain distributional assumptions about ψ. This is because of the two opposing effects that were explained in the first comparative static – there are less firms but the remaining firms are hiring more workers to cover the fixed cost.

2.2.5 Graphical example

Figure 2.1 shows a simple graphical example to explain the model. A firm that draws $\check{\psi}$ such that $\tilde{\psi}_j \leq \check{\psi} < \tilde{\psi}_{j+1}$ belongs to industry j. County $c+1$ spent a longer amount of time on the frontier than county c. As a result, the fixed cost associated with producing there is higher ($f(\psi_{c+1}) > f(\psi_c)$). The graph shows that there are fewer firms producing in counties that spent a longer time on the frontier. All firms that draw $\check{\psi} \geq \underline{\psi}_c$ produce in county c. However, all firms that draw $\check{\psi} < \underline{\psi}_{c+1}$ do not produce in county $c+1$. Consequently, firms in high "contractibility" industries sort into producing at locations that spent a longer time on the frontier.

Figure 2.1: Simple example of model



2.2.6 Empirical implications

Strictly speaking, the empirical analysis does not map perfectly to the conceptual framework. First, in the empirical analysis, I log transform the outcome variables to account for the skewness in their distributions. However, taking logs in the conceptual framework would result in the cross-derivatives becoming zero. Second, the conceptual framework makes a distinction between the number of firms locating in a county versus the number of firms producing in a county. In the empirical analysis, I do not make such a distinction. Nevertheless, the purpose of the conceptual framework is to demonstrate the sorting mechanism behind why certain firms and industries produce in certain counties.

2.3 Background and data

2.3.1 History: The amount of time that the county spent on the frontier

The idea of the "frontier" originated in the Census Bureau report "Progress of the Nation from 1890" by Porter et al. (1890). The report details the decade-by-decade westward migration and includes vivid maps of population density. In addition, the authors of the report argue that the frontier closed by 1890. Inspired by this report, Turner (1893) argue that the westward-moving frontier of settlement shaped early US history since it fostered a culture of "rugged individualism" – a term subsequently popularized by Hebert Hoover in his 1928 presidential campaign.

It is important to note that Turner's writings contain "departures from the historical record, overblown statements, and ethnocentric biases" (Bazzi et al. (2020)). In fact, Bazzi et al. (2020) cite writings to caution that the westward expansion involved land being violently taken from Native Americans. Far from the romantic notions of the frontier espoused by Turner, the frontier is actually associated with much violence.

I follow the definition used by Bazzi et al. (2020) to define frontier counties as those that (1) are in close proximity to the frontier line (100 km) and (2) have a population density below six people per square mile. Based on the replication data provided by Bazzi et al. (2020), the average time spent on the frontier is 14 years while the median is 13 years. The 25th percentile is 3 years and the 75th percentile is 21 years. The time spent on the frontier maps to ϕ_c in the conceptual framework.

2.3.2 Industries: Define industries based on how "contractible" inputs are

The measure of "contractibility" is based on the data from Nunn (2007b). The measure of how "contractible" an industry's inputs are is constructed in two steps. In the first step, inputs are classified based on the classification of goods by Rauch (1999): (1) goods that are sold on an organized exchange (e.g., oil), (2) goods that have a reference price (i.e., appear in catalogs) and (3) differentiated goods (i.e., goods that are not in the first two categories). Inputs in the first two categories have multiple buyers for the input. Therefore, Nunn (2007b) argues that the investments made by the suppliers of these inputs are not relationship-specific. If the buyer were to renegotiate a lower price ex post, the supplier could simply sell the input to another buyer.

In the second step, using the 1997 US Input-Output tables, for each output, the share of its inputs that are in the first two categories is calculated. This gives me the measure of "contractibility" that I use in this paper. Higher "contractibility" means that the inputs are less relationship-specific and are hence easier to specify in contracts. This measure of "contractible" is similar to Alfaro et al. (2019) and corresponds to one minus the measure in Nunn (2007b). Note that "contractibility" maps to ψ_j in the conceptual framework. Table 2.1 and Table 2.2 show the top twenty most and least "contractible" industries respectively.

2.3.3 Concordance and suppression of data

The outcome variables in this paper – the number of establishments and employment – come from the County Business Patterns (CBP) dataset. The data before 1998 are in terms of SIC while the data from 1998 onward are in terms of NAICS. There are substantial differences between SIC and NAICS. Therefore, to prevent concordance issues, I examine only years from 1998 onward. In addition, from 1998 to 2016, the NAICS changed its classification system three times so I map all industries to NAICS 1997.[6] When the mapping is not one-to-one, I apportion the number of establishments and employment equally across the multiple industries (simple average).

Finally, in the CBP data, employment for many county-industry cells is suppressed to preserve the confidentiality of the firm. For these cells, the CBP data reports employment as zero along with an "employment suppression flag". These flags take on 12 mutually exclusive ranges that contain

[6] In particular, from 2003-2007, I map NAICS 2002 to NAICS 1997. From 2008-2011, I map NAICS 2007 to NAICS 1997. Finally, from 2012-2016, I map NAICS 2012 to NAICS 1997.

the cell's true level of employment. To circumvent this issue, for the suppressed cells, I impute their employment as the midpoint of the range. For example, if a suppressed cell is listed as having employment being in the range of 500 to 999, I impute employment as 749.5. Finally, the last range (100,000 or more) has no upper bound. Therefore, for cells listed with this range, I set employment to 100,000.

Importantly, the CBP does not suppress the number of establishments in each county-industry cell. Therefore, while regressions involving employment as the outcome variable would suffer from measurement error, the regressions involving the number of establishments will not have this issue.

Table 2.1: Top twenty most "contractible" industries

NAICS97	Industry	Contractibility
311211	Flour Milling	0.968
324110	Petroleum Refineries	0.963
311221	Wet Corn Milling	0.960
311213	Malt Manufacturing	0.959
331315	Aluminum Sheet, Plate, And Foil Manufacturing	0.932
331312	Primary Aluminum Production	0.926
331316	Aluminum Extruded Product Manufacturing	0.912
325311	Nitrogenous Fertilizer Manufacturing	0.910
331319	Other Aluminum Rolling And Drawing	0.904
311212	Rice Milling	0.897
311615	Poultry Processing	0.891
324199	All Other Petroleum And Coal Products Manufacturing	0.868
331314	Secondary Smelting And Alloying Of Aluminum	0.868
327320	Ready-Mix Concrete Manufacturing	0.868
312210	Tobacco Stemming And Redrying	0.867
331419	Primary Smelting And Refining Of Nonferrous Metal (Except Copper And Aluminum)	0.859
311223	Other Oilseed Processing	0.855
325314	Fertilizer (Mixing Only) Manufacturing	0.854
325120	Industrial Gas Manufacturing	0.848
331492	Secondary Smelting, Refining, And Alloying Of Nonferrous Metal (Except Copper And Aluminum)	0.828

Table 2.2: Top twenty least "contractible" industries

NAICS97	Industry	Contractibility
336111	Automobile Manufacturing	0.014
336112	Light Truck And Utility Vehicle Manufacturing	0.014
334111	Electronic Computer Manufacturing	0.014
336120	Heavy Duty Truck Manufacturing	0.016
334515	Instrument Manufacturing For Measuring And Testing Electricity And Electrical Signals	0.037
334119	Other Computer Peripheral Equipment Manufacturing	0.043
334112	Computer Storage Device Manufacturing	0.043
334220	Radio, Television Broadcasting And Wireless Communications Eq. Mfg	0.044
334113	Computer Terminal Manufacturing	0.045
334511	Search, Detection, Navigation, Guidance, Aeronautical, And Nautical System And Instrument Manufacturing	0.046
334210	Telephone Apparatus Manufacturing	0.047
334517	Irradiation Apparatus Manufacturing	0.054
336411	Aircraft Manufacturing	0.059
334310	Audio And Video Equipment Manufacturing	0.063
336415	Guided Missile, Space Vehicle Propulsion Unit And Propulsion Unit Parts Mfg	0.082
336419	Other Guided Missile, Space Vehicle Parts And Auxiliary Equipment Mfg	0.082
336414	Guided Missile And Space Vehicle Manufacturing	0.086
339992	Musical Instrument Manufacturing	0.089
335314	Relay And Industrial Control Manufacturing	0.090
334510	Electromedical And Electrotherapeutic Apparatus Manufacturing	0.091

2.4 Outcomes at the county level and the industry level

I begin the empirical section of the paper by showing evidence of the first two comparative statics highlighted in the model.

2.4.1 Outcomes at the county level

To examine whether there are fewer establishments and lower employment in counties that spent a longer time on the frontier, I run the following regression:

$$Outcome_{ict} = \alpha_i + \delta_t + \beta Frontier_c + \gamma' X_c + \epsilon_{ict}$$

The outcome variables are the number of establishments and employment in industry i in county c in year t. There are a total of 3,104 counties and 19 years from 1998 to 2016. As a substantial number of these observations involve zeros in the outcome variable, I apply the inverse hyperbolic sine transform to the outcome variables. $Frontier_c$ is the number of decades that a county spent on the frontier and the data are from Bazzi et al. (2020).

Table 2.3: Relationship between time spent on the frontier and the number of establishments, as well as employment

	(1)	(2)	(3)	(4)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)				
Frontier	-0.025***	-0.017***	-0.022***	-0.024***
	(0.003)	(0.003)	(0.003)	(0.003)
R-squared	0.128	0.159	0.195	0.205
B. Outcome: Employment (Inverse hyperbolic sine transformed)				
Frontier	-0.070***	-0.049***	-0.063***	-0.068***
	(0.007)	(0.008)	(0.008)	(0.009)
R-squared	0.088	0.115	0.149	0.158
Observations	27,895,648	27,895,648	27,895,648	27,895,648
Geographical Controls		✓	✓	✓
Historical Controls			✓	✓
State FE				✓

Notes: All columns include industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

X_c is a vector of geographical and historical controls. The geographical controls are latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include the number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. α_i are industry fixed effects and δ_t are year fixed effects.[7] I cluster the standard errors at the county level.

Table 2.3 summarizes the relationship between time spent on the frontier and the number of establishments and employment in each county. Panel A shows the results for establishments while panel B shows the results for employment. In-line with the predictions of the model, the results suggest that there is a negative correlation between time spent on the frontier and the number of establishments, as well as employment in each county. In particular, every additional decade that a county spent on the frontier resulted in a 2.4% decrease in the (transformed) number of establishments and 6.8% decrease in the (transformed) number of workers.

2.4.2 Outcomes at the industry level

To examine if there are fewer establishments and lower employment in industries that are more "contractible", I run the following regression:

$$Outcome_{ict} = \kappa_c + \delta_t + \beta Contractibility_i + \gamma' X_i + \epsilon_{ict}$$

$Contractibility_i$ measures how "contractible" industry's i inputs are. This measure of "contractibility" is based on the 1997 US Input-Output Use tables. Higher "contractibility" means that the inputs are less relationship-specific and hence easier to specify in contracts. κ_c denotes county fixed effects. X_i consists of industry controls such as the log value add of the industry in 1990 and its TFP in 1990. I cluster the standard errors at the industry level.

[7]These variables are from the replication data set provided by Bazzi et al. (2020). More details on the various data sources can be found in Appendix K of their paper.

Table 2.4: Relationship between "contractibility" and the number of establishments, as well as employment

	(1)	(2)	(3)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)			
Contractibility	-0.026	-0.010	-0.009
	(0.024)	(0.023)	(0.052)
R-squared	0.236	0.251	0.275
B. Outcome: Employment (Inverse hyperbolic sine transformed)			
Contractibility	-0.064	-0.012	-0.017
	(0.066)	(0.065)	(0.138)
R-squared	0.190	0.203	0.220
Observations	27,895,648	27,659,744	27,659,744
Industry Controls		✓	✓
3-digit Industry FE			✓

Notes: All columns include county and year fixed effects. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at industry level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 2.4 summarizes the relationship between "contractibility" and the number of establishments and employment in each industry. Panel A shows the results for establishments while panel B shows the results for employment. In-line with the predictions of the model, the results suggest that there is a negative correlation between "contractibility" and the number of establishments, as well as employment in each industry. However, the results are imprecisely estimated. One thing to note is that the number of observations falls from 27,895,648 in column 1 to 27,659,744 in column 2. This is because the NBER-CES data, which is where I obtained the data for the industry controls, do not have data for four industries in 1990.[8]

[8] These industries along with their NAICS 1997 codes are "Tire retreading" (326212), "Retail bakeries" (311811), "Software reproducing" (334611) and "Dental laboratories" (339116).

2.5 Sorting of industries to counties

Next, using a difference-in-differences (DiD) empirical strategy, I examine whether high "contractibility" industries sort to counties that spent a longer time on the frontier:

$$Outcome_{ict} = \alpha_i + \kappa_c + \delta_t + \beta Contractibility_i \times Frontier_c + \gamma' X_{ic} + \epsilon_{ict} \qquad (2.2)$$

The variables are the same as what was previously defined. As a substantial number of these observations involve zeros in the outcome variable, I apply the inverse hyperbolic sine transform to the outcome variables. I cluster the standard errors at the county level. A positive effect (i.e., $\beta > 0$) would suggest that high "contractibility" industries sort to counties that spent a longer time on the frontier. This maps to the third comparative statics of the model.

This DiD strategy exploits two cross-sectional variations. The first cross-sectional variation comes from the fact that different counties spent different number of years on the American frontier. The second cross-sectional variation arises because different industries have different levels of "contractibility". For example, industries that predominantly use inputs that are sold on an organized exchange (e.g., oil) are highly "contractible". This is because the inputs that they use are less relationship-specific and hence easier to specify in contracts.

2.5.1 Results and discussion

Table 2.5 reports the results as to whether time spent on the frontier affects the present-day distribution of manufacturing industries. Panel A reports the results for the number of establishments while panel B reports the results for employment.

In column 1 which controls for industry, county and year fixed effects, the estimated coefficient for $Contractibility_i \times Frontier_c$ is positive and statistically significant for both the establishments and employment regressions. This suggests that high "contractibility" industries sort to counties that spent a longer time on the frontier.

Table 2.5: Effect of time spent on the frontier on the present-day composition of manufacturing industries

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Outcome: Establishments (Inverse hyperbolic sine transformed)					
Contractibility X Frontier	0.014***	0.009***	0.011***	0.008***	0.009***	0.015***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
R-squared	0.358	0.358	0.359	0.358	0.422	0.490
	B. Outcome: Employment (Inverse hyperbolic sine transformed)					
Contractibility X Frontier	0.036***	0.020***	0.027***	0.019***	0.021***	0.043***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.009)
R-squared	0.274	0.274	0.274	0.272	0.318	0.370
Observations	27,895,648	27,895,648	27,895,648	27,659,744	27,659,744	27,659,744
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Nonetheless, there could be concerns that there are omitted variables at the industry-county level that are correlated with $Contractibility_i \times Frontier_c$ and the outcome variables. For example, geography could determine the amount of time that a county spent on the frontier. A county that is surrounded by water bodies and mountains tend to spend more time on the frontier as these geographical constraints make it difficult for settlers to explore beyond the county. At the same time, these geographical features also determine the location of industries. For example, industries that mainly use natural resources are more "contractible" and tend to be located in areas where there is an abundance of these resources. Therefore, to address these concerns, in column 2, I include a series of predetermined county-level **geographical controls** interacted with $Contractibility_i$. These include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Reassuringly, the estimated effect remains robust to the inclusion of these controls.

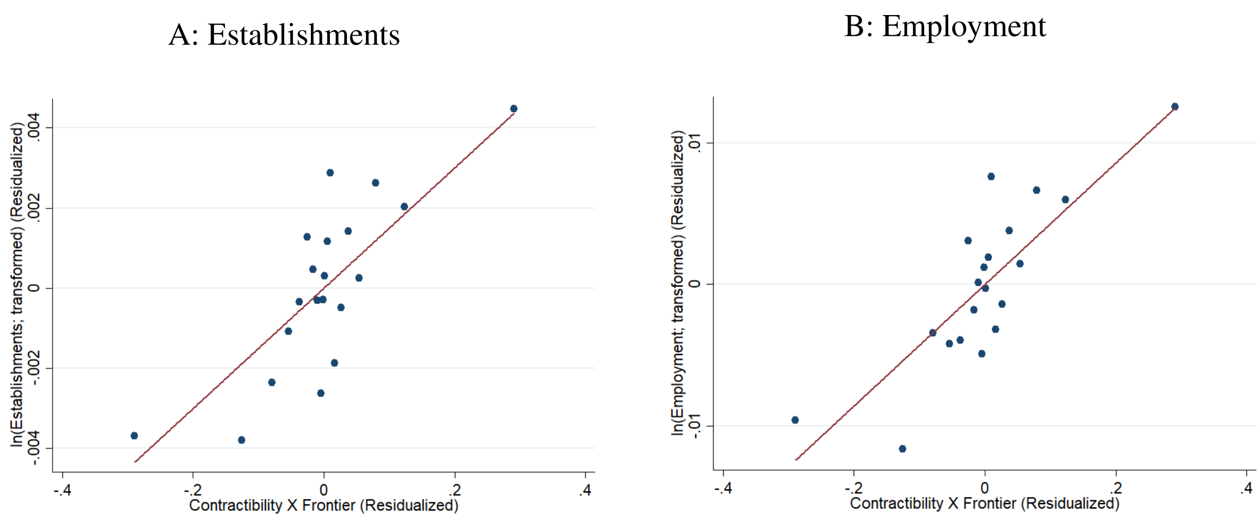
However, there could be concerns that there are other historical events that affect the sorting of industries to counties. To address this potential issue, in column 3, I augment the geographical controls with a series of predetermined county-level **historical controls** and interact them with $Contractibility_i$. The historical controls include the number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Column 3 shows that the estimated coefficient remains stable to these controls.

In column 4, to address concerns that other industry characteristics might be driving the sorting pattern, I include **industry controls** interacted with $Frontier_c$. These industry controls are the log value add of the industry in 1990 and its TFP in 1990. Finally, I subject the results to extremely demanding interacted fixed effects that leave very limited identifying variation. In column 5, I include state-by-industry fixed effects which means that the identifying variation comes from industries within each state. In column 6, I add county-by-3-digit industry fixed effects. Doing so restricts the identifying variation to within each 3-digit industry at the county level. These interacted fixed effects would also help to control for some of the measurement error that arises in the mapping of NAICS industries across the different years, as well as from using the midpoint to impute employment in the suppressed cells.

Despite the inclusion of these extremely demanding fixed effects, the estimated coefficient remains remarkably stable and highly significant, suggesting that high "contractibility" industries do indeed sort to counties that spent a longer time on the frontier. In particular, every additional decade that a county spent on the frontier resulted in a 1.4% increase in the (transformed) number of establishments in the highest "contractible" industry (0.968) as compared to the lowest "contractible" industry (0.014) in that county. As for employment, every additional decade that a county spent on the frontier resulted in a 4.1% increase in the (transformed) number of workers employed in the highest "contractible" industry (0.968) as compared to the lowest "contractible" industry (0.014) in that county. The results in Table 2.5 are consistent with the mechanisms proposed in the conceptual framework – only the high "contractibility" firms are able to pay the high fixed cost associated with producing in counties that spent a longer time on the frontier.

To check that the results are not driven by particular outliers, I plot a binned scatter plot of the residues of the outcome variables against the residues of $Contractibility_i \times Frontier_c$ based on the results in column 6. Figure 2.2 presents the scatter plots and shows that the estimated effects would be larger if the outliers are excluded.

Figure 2.2: Binned scatter plot of the residues



Instead of “contractibility”, Levchenko (2007) uses an alternative measure of holdup – the Herfindahl index of intermediate input use. The Herfindahl index of intermediate input use increases with concentration. Levchenko (2007) argues that industries that rely on a large number of suppliers to provide small quantities of many inputs are more vulnerable to holdup problems. Therefore, a larger Herfindahl index implies less holdup.

Table 2.6 shows the results where instead of “contractibility”, I use the Herfindahl index of intermediate input use based on the 1997 Input-Output Use tables. The results using this alternative measure of holdup are slightly larger compared to the baseline results.

Table 2.6: Using Herfindahl index of intermediate input use instead of “contractibility”

	(1)	(2)	(3)	(4)	(5)	(6)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)						
Herfindahl X Frontier	0.054***	0.034***	0.044***	0.036***	0.042***	0.026***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.005)
R-squared	0.358	0.359	0.360	0.358	0.422	0.490
B. Outcome: Employment (Inverse hyperbolic sine transformed)						
Herfindahl X Frontier	0.120***	0.053***	0.083***	0.064***	0.084***	0.049***
	(0.014)	(0.016)	(0.017)	(0.016)	(0.017)	(0.014)
R-squared	0.274	0.274	0.275	0.273	0.318	0.370
Observations	27,895,648	27,895,648	27,895,648	27,659,744	27,659,744	27,659,744
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

2.5.2 Robustness checks

Next, I subject the results to a series of robustness checks. The robustness checks can be broadly classified into three categories. First, subjecting the results to different sub-samples or adding more controls. Second, checking that the results are similar even when accounting for the zeros in the outcome variable. Finally, accounting for spatial correlations.

Doing the analysis separately for coastal and non-coastal counties. There could be concerns that the sorting pattern could be due to geography and not "contractibility". This is unlikely to be the case since the regressions already control for geographical factors interacted with "contractibility". These geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Nevertheless, I take further steps to rule out the role of geography by splitting the sample into coastal and non-coastal counties. This is important because Rappaport and Sachs (2003) argue that economic activities are overwhelmingly concentrated at coastal areas. If we see the sorting pattern only in the coastal sample but not in the non-coastal sample, then we would worry that it is the effect of being a coastal county that is driving this sorting pattern. However, if we continue to see the same sorting pattern in the non-coastal counties, then it is unlikely that geography is the main driver of the result. Tables B1 and B2 report the results for non-coastal and coastal counties respectively. Reassuringly, in the non-coastal sample, the coefficient estimates continue to be positive and highly significant. However, in the coastal sample, the coefficient estimates while positive, become imprecisely estimated once state-by-industry fixed effects are added. This suggests that being a coastal county actually acts against the pattern of high "contractibility" industries sorting to counties that spent a longer time on the frontier.

Including controls for population density. Since the definition of the frontier is based on population density, there could be worries that the sorting pattern is just picking up the effects of density and not the time spent on the frontier. To address this concern, I include population density variables as controls. These controls are the county's population density in 1910 and the number of years between 1790 and 1890 where the county had a low population density (defined as less than six people per square mile). The results in Table B3 show that the coefficient estimates are statistically and economically important even after controlling for population density. This suggests that there are aspects of time spent on the frontier that cannot be attributed to density, supporting the argument that time spent on the frontier affects the present-day sorting pattern of industries to counties.

Dropping observations with zeros in the outcome variable. The baseline regression uses a total of 27,895,648 observations. Of these, around 25,288,000 observations have zeros in the outcome variables. This could lead to concerns that the results are being overwhelmingly driven by observations with zeros in the outcome variables. To address this concern, I drop all the observations with zeros in the outcome variables and re-run the regressions. Table B4 shows that when these observations are dropped, in all the columns, the estimated coefficient remains positive and statistically significant but are now larger. This suggests that the 25,288,000 observations that have zeros in the outcome variables are causing the estimated coefficients to be attenuated towards zero.

Using Poisson pseudo-maximum likelihood (PPML) regression to account for the zeros in the outcome variable. Instead of dropping observations or applying the inverse hyperbolic sine transform, another method to account for zeros in the outcome variable is to use a PPML regression. This approach is consistent with the extensive literature – for example, Santos Silva and Tenreyro (2006) and Head and Mayer (2013) – that shows that PPML is a consistent and relatively efficient estimator for specifications featuring a large number of zeros in the data. Table B5 contains the results using the PPML regression. Compared to the baseline DiD results, the magnitudes of the estimated coefficients are now six to ten times larger. This suggests that the baseline results are a lower bound.

Dropping counties that changed names or FIPS codes, as well as counties that had substantial boundary changes between 1980 and 2016. County boundaries often change over time due to name changes, annexations and incorporations. Therefore, when calculating the total time that each county spent on the frontier, Bazzi et al. (2020) maintain consistent units of observations over time by harmonizing boundaries to 2010 county boundaries.[9] This is the same approach that is used in Hornbeck (2010). While total time spent on the frontier has been harmonized to 2010 county boundaries, the outcome variables (number of establishments and employment) from the CBP datasets have not been harmonized to time consistent boundaries. According to the Census Bureau website, between 1980 and 2016, there were a total of 67 counties that changed names or FIPS codes and/or had substantial boundary changes.[10] This means that these 67 counties do not have consistent boundaries as the 2010 boundaries used by Bazzi et al. (2020) in their calculation of total time spent on the frontier. Therefore, to check if the results are sensitive to boundary changes, I drop these 67 counties and re-run regression 2.2. Table B6 shows that the estimated coefficients remain robust to dropping these 67 counties.

Dropping counties which spent zero years on the frontier. Bazzi et al. (2020) code 608 counties as having spent zero years on the frontier. These counties fall into two categories. First, since the Census only began in 1790, counties within 100 km of the frontier line and already have a population density above six people per square mile in the 1790 Census are coded as having spent zero years on the frontier. Second, since the end-point of the frontier is 1890, counties beyond the frontier but have low population density in 1890 are nevertheless coded as spending zero years on the frontier. To check if the results are sensitive to these measurement errors, I re-run regression 2.2 without these 608 counties. Table B7 shows that the results remain robust to this check.

[9]Bazzi et al. (2020) show that the location of the frontier is very similar when using contemporaneous historical boundaries.

[10]The US Census Bureau defines substantial county boundary changes as those affecting an estimated population of 200 or more; changes of at least one square mile where an estimated population number was not available, but research indicated that 200 or more people may have been affected; and annexations of unpopulated territory of at least 10 square miles.

Dropping industries which did not have unique mapping when industry classifications changed across the years. Between 1998 and 2016 (the period of my study), the industry classification system changed three times – 2002, 2007 and 2012. When the mapping between the different editions of the NAICS is not one-to-one, I apportion the number of establishments and employment equally across the multiple industries (i.e., simple average). Therefore, there could be concerns that the results are driven by this particular method of dealing with mappings that are not one-to-one. To address this concern, I drop all industries where the mapping across the different NAICS years is not one-to-one. Out of the 474 NAICS 1997 industries in the original regression, only 298 of them have a one-to-one mapping across all the different NAICS years. Table B8 shows that the results remain robust even when using only these 298 industries.

Accounting for spatial correlation by clustering standard errors based on 60-square mile grid cells. Table B9 shows that the results remain robust even when using this method of clustering the standard errors to account for spatial correlations.

2.6 Potential mechanisms behind the sorting pattern

In this section, I explore various potential mechanisms that could explain why high "contractibility" industries sort to counties that spent a longer time on the frontier.

2.6.1 First-mover advantage of industries

One potential explanation behind the sorting pattern is that of first-mover advantage that persists due to path dependency. Under this explanation, the present-day high "contractibility" industries that sort to counties that spent a longer time on the frontier are the same industries that sort to these counties in the past. Due to path dependency, this sorting pattern continues to persist till today even if the initial reasons for why these industries locate in these counties are no longer relevant. In other words, if a different set of industries had sorted themselves to these counties, then that would be the set of industries that we see today.[11]

To test the first-mover advantage mechanism, I do two things. First, I examine whether high "contractibility" industries of the past sort to counties that spent a longer time on the frontier. Second, I check if the high "contractibility" industries of the past are the same high "contractibility" industries today.

Historical data and defining "contractible" industries of the past

The present-day measures of "contractibility" are based on the 1997 US Input-Output tables. However, IO tables do not exist for the time period 1790 to 1890 (period when the country was expanding westward). Therefore, to proxy for the "contractibility" of industries, I turn to the data from the Census of Manufacturing (1860 to 1880) which were recently digitized by Hornbeck and Rotemberg (2019). The data contain information at the county-industry level for the number of establishments, employment, value of capital invested, labor cost, material cost and production value. Over 1,000 industries are listed in the raw data but Hornbeck and Rotemberg (2019) group these into 159 categories.

To construct a proxy measure for the "contractibility" of industries, I use the share of capital in the total cost of production (i.e., value of capital invested divided by the sum of the value of capital invested, labor cost and material cost). The idea behind this simple proxy of "contractibility" is that labor and material inputs are more susceptible to holdup whereas the capital stock is not. This is because the capital stock accumulates over time and is sunk. Therefore, since the capital stock is sunk, in industries where the share of the capital stock is higher, there is less scope for holdup. Nevertheless, it is important to note that this simple measure of holdup is more likely to hold historically than in present times as inputs were relatively more homogeneous and production processes were less complex in the past – the more processed the good, the harder it is to specify it

[11] This explanation assumes the presence of multiple equilibria.

in contracts.

Examining if the sorting pattern also happened in the past when the country was expanding

To examine this, I run the same regression as regression 2.2: $Outcome_{ict} = \alpha_i + \kappa_c + \delta_t + \beta Contractibility_i \times Frontier_c + \gamma' X_{ic} + \epsilon_{ict}$. The two differences are that first, instead of the treatment variable being $Contractibility_i \times Frontier_c$, I use the capital share in the total cost of production in 1860 as the proxy for historical "contractibility". To avoid the issue of reverse causality, I use the 1860 data to construct the share of capital in the total cost of production and use data from 1870 and 1880 for the outcome variables (number of establishment and employment). The second difference is that the controls are now based on 1860 data. For example, the number of years that the county was connected to the railroad by 1860, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1860.

The westward expansion of the country meant that county boundaries were often evolving. Therefore, to ensure that counties are comparable across time, I do three things. First, I use the crosswalk from Eckert et al. (2020) to aggregate all historical variables in terms of county boundaries in 2010. Second, I use only counties that appeared in both 1870 and 1880. Finally, I use only counties that exited the frontier by 1870. Table 2.7 presents the results and suggests that the high "contractibility" industries of the past did indeed sort to counties that spent a longer time on the frontier. However, when county-by-broad industry fixed effects are added, the coefficient estimates while positive are now imprecisely estimated.

Examining if the high "contractibility" industries of the past are the same high "contractibility" industries today

While the previous subsection showed that the high "contractibility" industries of the past did indeed sort to counties that spent a longer time on the frontier, it could be the case that these industries are not the same as the high "contractibility" industries today. To test this, I first map the 159 industries from the Census of Manufactures (1860 to 1880) to the IND1950 industry classifications used in the IPUMS complete count Census microdata. There is no publicly available crosswalk for this so I constructed my own. Next, I map the 474 industries from the CBP data (which uses NAICS) to the IND1950 industry classifications. The crosswalk for this is based on information provided by IPUMS. As a result of these two steps, I am able to map the industries of the past and the present to a common set of industries based on the IND1950 classifications.

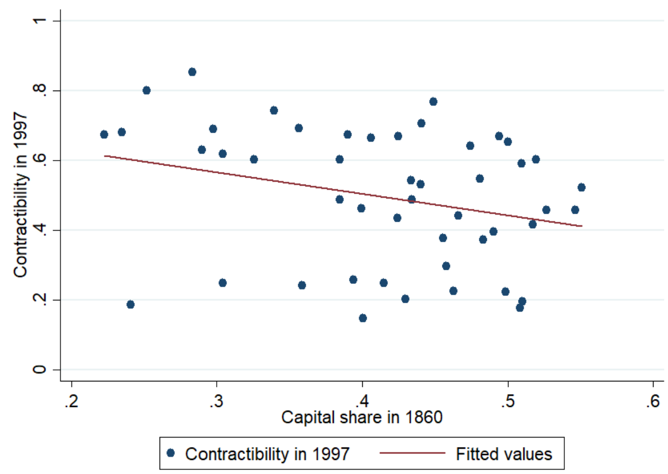
Table 2.7: Effect of time spent on the frontier on the historical composition of manufacturing industries

	(1)	(2)	(3)	(4)	(5)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)					
Share of capital in 1860 X Frontier	0.037***	0.033***	0.031***	0.030***	0.005
	(0.004)	(0.005)	(0.004)	(0.005)	(0.003)
R-squared	0.326	0.326	0.326	0.497	0.669
B. Outcome: Employment (Inverse hyperbolic sine transformed)					
Share of capital in 1860 X Frontier	0.043***	0.045***	0.044***	0.042***	0.006
	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)
R-squared	0.299	0.299	0.299	0.436	0.587
Observations	583,702	583,702	582,664	582,664	582,664
Contractibility X Geographical Controls		✓	✓	✓	✓
Contractibility X Historical Controls			✓	✓	✓
State X Industy FE				✓	✓
County X Broad Industry FE					✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1860, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1860. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Following this, I plot the industry's "contractibility" in 1997 against its share of capital in the total cost of production in 1860. Table 2.3 presents this scatter plot and it shows a negative relationship. This suggests that the high "contractibility" industries of the past are different from the high "contractibility" industries today.

Figure 2.3: Relationship between past and present "contractibility"



In fact, there is evidence to suggest that in present times, the high "contractibility" industries of the past actually **sort away** from counties that spent a longer time on the frontier. To see this, I map all present-day NAICS industries from the CBP data to IND1950 and re-run regression 2.2 using the historical measure of "contractibility". Table 2.8 shows the results of this regression. In panel A which presents the results for the number of establishments, across all the columns, the estimated effects are negative. This suggests that the high "contractibility" industries of the past actually **sort away** from counties that spent a longer time on the frontier. The results for employment in panel B are less conclusive as the coefficient estimates although negative in the columns with the most stringent controls, are imprecisely estimated.

The results using historical data (1880 to 1890) in Table 2.7 being positive while the results using modern data (1998 to 2016) Table 2.8 being negative suggests two things. First, it shows that in historical times, the high "contractibility" industries of the past sort to counties that spent a longer time on the frontier. However, in present times, these high "contractibility" industries of the past actually **sort away** from counties that spent a longer time on the frontier. Second, the industries of the past that sort to counties that spent a longer time on the frontier are different from the present-day industries that sort to these counties.

Table 2.8: Effect of time spent on the frontier on the present-day composition of manufacturing industries (historical measure of "contractibility")

	(1)	(2)	(3)	(4)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)				
Capital Share 1860 X Frontier	-0.039***	-0.038**	-0.066***	-0.075***
	(0.014)	(0.016)	(0.016)	(0.018)
R-squared	0.602	0.603	0.604	0.663
B. Outcome: Employment (Inverse hyperbolic sine transformed)				
Capital Share 1860 X Frontier	0.004	0.018	-0.012	-0.022
	(0.037)	(0.043)	(0.043)	(0.046)
R-squared	0.498	0.499	0.500	0.546
Observations	2,830,848	2,830,848	2,830,848	2,830,848
Contractibility X Geographical Controls		✓	✓	✓
Contractibility X Historical Controls			✓	✓
State X Industy FE				✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Standard errors clustered at county level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

2.6.2 Culture and institutions

Another explanation could be that the time spent on the frontier has both contemporaneous and persistent effects on the counties' culture and institutions. Consequently, this affects the type of industries that sort to the various counties. In a ground-breaking paper, Bazzi et al. (2020) provide evidence that the frontier attracted individualistic people. In addition, the uncharted physical environment at the frontier with little social infrastructure to turn to made its residents even more individualistic over time.

Remarkably, they also show that the culture of "rugged individualism" continues to persist even after the frontier had closed. For example, in the mid-20th century, infrequent names given to children continued to be more pervasive in counties that spent a longer time on the frontier. In addition, they show that even in the late 20th century, residents of these counties exhibit more pervasive individualism, prefer less redistribution, lower public spending and less social protection in terms of minimum wages, gun control and environment protection. What their results show us is that the individualistic culture and institutions that developed in the early days became fundamental to a place over time and hence persist till today.

How then does the "rugged individualism" of a county affect the sorting of industries? The intuition behind this is that in counties with more individualistic culture and institutions, individuals are less likely to trust other people. Therefore, anything that is not "contractible" becomes harder and more costly to enforce – consistent with the assumption in the model that there is a higher cost when producing at counties that spent a longer time on the frontier. Consequently, only the more "contractible" industries locate in counties that spent a longer time on the frontier. This argument is consistent with why we observe the high "contractibility" industries of the past and present sorting to counties that spent a longer time on the frontier – although the high "contractibility" industries of the past are different from the present-day high "contractibility" industries.

2.7 Conclusion

What explains the spatial distribution of economic activities? In this paper, I go beyond the commonly studied factors and instead examine the role that a particular episode of history – time spent on the frontier – plays in explaining the spatial distribution of manufacturing industries in the US. First, I find that there are fewer establishments and lower employment in counties that spent a longer time on the frontier. The same results hold for industries that are more "contractible" (i.e., easier to specify in contracts and hence less susceptible to holdup). Second, using a difference-in-differences strategy, I find that firms in high "contractibility" industries sort into producing at counties that spent a longer time on the frontier. Relying on the arguments in Bazzi et al. (2020), I hypothesize that due to "rugged individualism", individuals in counties that spent a longer time on the frontier are less likely to trust other people. Therefore, anything that is not "contractible" becomes harder and more costly to enforce. Consequently, only the more "contractible" industries locate in counties that spent a longer time on the frontier.

Chapter 3: The Inspector Calls: The Effect of Local Land Use Regulations and NIMBY-ism on Housing Prices (with Di Song Tan)

3.1 Introduction

"In the past 25 years, construction has come to face enormous challenges from any local opposition. In some areas it feels as if every neighbor has veto rights over every project...To most residents, a new project is nothing but a bother. They don't care about the welfare received by the new resident, or the benefits earned by the builders or by the employers who have to pay lower wages when housing costs are lower." - Glaeser (2014)[1]

Land use regulations set guidelines and rules that control the supply of land that can be used for housing, retail, commerce and industry. These regulations often take the form of restrictions on land use (e.g., zoning law), the density of development (e.g., building-height limits and minimum lot size requirements) and the volume of development (e.g., urban growth boundaries and caps on building permits). Today, with more than half of the world's population living in urban areas, land use regulations have become common-place.

Traditionally, in most countries, land use regulations are determined by national planning agencies. The US is one of the few exceptions where land use regulations are mainly controlled by local governments.[2] Gyourko and Molloy (2015) note that the initial land use regulation in the United States (US) began with the intention of separating different types of land use so as to limit negative

[1] https://www.cato.org/publications/cato-online-forum/land-use-restrictions-other-barriers-growth

[2] Gyourko and Molloy (2015) note that this is because the US Constitution did not grant the federal government authority to regulate land. As a result, states have generally vested this power with local governments.

externalities. However, Cheshire et al. (2014) note that in recent times, countries such as the United Kingdom (UK) are increasingly moving towards localism. This has led to increasing concerns that land use regulations are being used by locals to engage in NIMBY ("not in my back yard") behavior.

Against the backdrop of these developments, this paper examines how land use regulations and NIMBY-ism affect housing prices in the UK. For the purpose of our paper, we define NIMBY-ism as local power. In particular, it is the actions taken by individuals to object to the siting of developments that are perceived as unpleasant in their own neighborhood, while raising no such objections to similar developments elsewhere.[3] Studying the UK's experience is important because increasingly, many countries are adopting restrictions that are similar to the UK's containment policies. Our analysis consists of the following steps.

First, we model the application and approval process as a multi-stage sequential game between the developers, local planning authorities and the inspectors. In the UK, individuals and developers have to apply to the local planning authority to seek development permission (ownership alone does not confer the right to develop the land). Applicants who have their plans rejected can appeal to the Secretary of State, via the Planning Inspectorate. The Planning Inspectorate then assigns an inspector to decide whether to overturn the local authority's decision. We follow the institutional context closely in our model. In particular, in our model, the developer chooses the intensity of residential development. The local authority then accepts or rejects the project based on the level of NIMBY-ism in the area. Our model yields two key propositions – (i) there is negative selection into appeals because the developments that are appealed are the ones which have a greater negative impact on amenities. In addition, (ii) local power exacerbates selection into appeals. In locations with high levels of NIMBY-ism, even relatively benign projects end up being appealed. This is because developers have to give up too much of their profits to counter NIMBY-ism and so they are better off getting their plans rejected by the local authority and gambling on drawing an inspector

[3]We adapt this definition from the Oxford English Dictionary.

who is less sympathetic towards locals' NIMBY behavior.

This analysis suggests that we can understand the prevalence of NIMBY-ism from the selection of projects into our appeals sample. If only the worst projects, which have a deleterious impact on the neighborhood, select into appeals, then NIMBY-ism is not so serious; the planning process is working as intended as bad apples are rejected by the planners. On the other hand, if fairly benign projects select into appeals, then NIMBYs may be prevailing.

Second, we embark on our reduced form analysis as informed by our model. We begin by examining the effect of a successful appeal (overturning the local authority's decision) on housing prices. To establish causality, we employ two research designs: (i) an OLS using a difference-in-differences (DiD) approach; and (ii) an instrumental variables (IV) approach. The latter exploits the fact that inspectors who are assigned to the appeals have different preferences and are quasi-randomly assigned to the cases. This means that we can use the leniency of the inspectors as an instrument for whether an appeal is successful.[4] We find that overturning the local authority's decision does not lead to a large fall in housing prices. In fact, our IV estimates suggest a positive impact of overturning decisions. This suggests that NIMBY-ism may be fairly prevalent across the UK.

Our paper contributes to the burgeoning literature on the effects of land use regulations in two ways. First, the impact of land use regulations on housing prices have been documented in many empirical studies such as Albouy and Ehrlich (2018), Shertzer et al. (2018), Hilber and Vermeulen (2016), Turner et al. (2014), Libecap and Lueck (2011), Saiz (2010), Glaeser et al. (2005), Quigley and Raphael (2005) and Cheshire and Sheppard (2002). In addition, with the exception of Hilber and

[4] The use of expert leniency as an instrumental variable has been used widely in the economics literature. For example, the leniency of judges has been used to estimate the impact of eviction on poverty (Humphries et al. (2018)), incarceration on economic and family outcomes (Aizer and Doyle Jr (2015); Bhuller et al. (2020)), the effect of pretrial detention on legal and economic outcomes (Dobbie et al. (2018)), the effect of foster care on child outcomes (Doyle (2008)), the effect of disability on labor supply (Dahl et al. (2014); Kostøl et al. (2019)), and even the effect of patents on innovation (Galasso and Schankerman (2015)).

Vermeulen (2016) and Cheshire and Sheppard (2002), most of these studies are in the context of the US and not the UK. To the best of our knowledge, our paper is one of the first few to provide causal estimates of the effect of land use regulations in the UK. Our paper is also related to the literature on the effect of land use regulations on welfare. For example, Turner et al. (2014) and Hsieh and Moretti (2019) show that land use regulations lead to large decreases in welfare.

Second, and more importantly, our paper demonstrates why it is not correct to view land use regulations as simply being a supply shock.[5] Before Turner et al. (2014), the literature often took contradictory positions as to whether increases in housing prices indicate welfare increases or decreases. One position was that price increases reflected welfare improvements. This view is predicated on viewing land use regulations as improving the amenities of an area and hence a demand shock. The other position was that price increases reflected welfare lost. In this framework, land use regulations are viewed as a supply shock. Following Turner et al. (2014), we model land use regulations as being both a demand and supply shock. We then show how land use regulations being both a demand and supply shock have political economy implications.

In terms of the political economy literature, to the best of our knowledge, we contribute by being one of the first to empirically quantify how NIMBY-ism affects housing prices. In recent times, there has been a growing number of papers such as Parkhomenko (2020) and Ortalo-Magné and Prat (2014) that formalize Fischel's homevoter hypothesis by modelling land use regulations as a function of political competition between renters and owners. Besides the role of homeowners in the local political process, Calabrese et al. (2007), Epple et al. (1988) and Hamilton (1975) argue that land use regulations are ways to prevent certain types of households from entering a community. Similar to these papers, our paper models land use regulations as a political economy

[5] Gyourko and Molloy (2015) note that the predicted effects of regulations become less obvious if households can move freely among cities. This is because under spatial equilibrium, utility must be equal across cities. Therefore, population flows can erode away price differences across locations. The price differences across areas thus reflect the amenity value of growth controls and not the lower elasticity of housing supply.

mechanism to protect the interest of particular local groups ("insiders") at the expense of other groups ("outsiders"). In addition, just like Hilber and Robert-Nicoud (2013) and Glaeser et al. (2005), we also show how developers react to the local political process. Most of these political economy papers are mainly theoretical papers or calibration exercises which do not quantify the effect of NIMBY-ism. Our paper therefore contributes to this literature by providing an empirical estimate of how NIMBY-ism affects housing prices.

More broadly, our paper is related to the literature which examines the effects of housing policies. Examples of housing policies other than land use regulations include, urban revitalization (Greenstone and Gallagher (2008), Rossi-Hansberg et al. (2010)), direct provision of housing (Collins and Shester (2013)), price controls (Autor et al. (2014), Diamond et al. (2019)), and taxes and subsidies (Baum-Snow and Marion (2009), Collinson and Ganong (2018), Diamond and McQuade (2018)). Since land use regulations are a form of place-based policies, our paper relates to the literature on the use of place-based polices to rectify regional disparities. Kline and Moretti (2014) provide a good overview on the economics of place-based policies. In addition, because land use regulations limit the size of cities, our paper is also related to the literature on optimal city size. Au and Henderson (2006) find that migration restrictions have resulted in many undersized cities in China. To the extent that there are huge benefits from urban agglomeration, the costs of being undersized due to land use regulations are potentially high.

Finally, our paper is related to the literature on the economic effects of the decentralization of policy responsibilities to local governments. A review by Pike et al. (2012) suggests that there is no clear relationship between decentralization and economic outcomes. For example, Zhang and Zou (1998) find that decentralization is associated with lower economic growth at the provincial level in China. However, Akai and Sakata (2002) and Stansel (2005) find a positive relationship between decentralization and economic growth in the US. Yet another group of studies such as Davoodi and Zou (1998), Xie et al. (1999) and Rodríguez-Pose and Bwire (2004) find that there is no link between decentralization and economic outcomes. This literature is somewhat dated and does not establish causality. By examining how the decisions of local governments are exogenously overturned, our paper presents an attempt at reviving this literature.

The rest of the paper proceeds as follows. Section 3.2 provides the background to land use regulations in the UK and the rise of local power. Section 3.3 develops the theoretical model which we use to interpret our reduced form regressions. Section 3.4 describes the data sources which we use for our analysis. Section 3.5 explains our empirical strategy and Section 3.6 presents the results. Finally, Section 3.7 concludes.

3.2 Background to Land Use Regulations in the UK

3.2.1 Overview

Land use regulation in the UK has a long and storied history – Corkindale (1999) notes that as early as 1540, Queen Elizabeth I forbade any new buildings within three miles of the City gates of London. The present-day land use regulations in the UK has its roots in the Town and Country Planning Act which came into law in 1947. The purpose of the 1947 act was to contain urban areas and stop them from spilling out into the surrounding countryside as well as preserve amenities of various kinds. The act also separated land uses which might be incompatible (e.g., industry from residential). At the same time, the act also aimed to provide lower density and greener living conditions in the new towns.

In order to do so, the act sets out the principle of "development control" which means that planning permission is required for land development – ownership alone no longer conferred the right to develop the land. As a result, owners have to apply and seek permission from their local planning authorities whenever they wanted to develop their land. The local planning authorities would only approve the development if it is consistent with the local development plan.[6] The UK system therefore controls the amount of land available not just for housing but for all urban uses of land including offices, retail and commercial.

There have been many modifications since 1947.[7] For example, in the 1950s, the restrictions were tightened as "Greenbelt" boundaries were established. This resulted in increasing amount of land around towns and cities being taken out of the effective land supply. Later revisions of the act were legislated in 1962, 1971 and 1990. While the 1990 act is the current legislation, the act has been substantially amended and added to, for example, through the use of legislative orders. Nonetheless, the 1947 act established an approach and framework that has not been superseded.

Today, the main planning acts that are in force in England are the Town and Country Planning Act (1990), Planning and Compulsory Purchase Act (2004), Planning Act (2008) and Localism Act (2011). These acts are also read alongside a series of planning policy documents and guidelines such as the National Planning Policy Framework which was published in March 2012. Other national policies and restrictions such as "Greenbelts", "Sites of Special Scientific Interest" and "Areas of Outstanding Natural Beauty" also form part of the planning regulations. Many of the recent reforms such as the Localism Act (2011) was aimed at giving locals more power in deciding land use in their neighborhood. For example, neighborhood plans were introduced by the Localism Act (2011). These gave communities the direct power to develop a shared vision for their neighborhood and shape the development and growth of their local area. The neighborhood plans are extremely powerful. This is because when considering whether to grant planning permission, the local authorities have to ensure that the proposed development is consistent with the neighborhood plans.

[6] Today, many parts of England have three tiers of local government – (i) county council, (ii) district, borough or city councils and (iii) parish or town councils. The district councils are the ones that are typically responsible for most planning matters including preparing local development plans and approving planning applications.

[7] Each country within the UK has its own variations in their planning system. Since the English system is the dominant one, the details which we provide here are based on the system in England.

A planning application needs to be submitted to the local authorities if an individual or developer wants to (i) build something new, (ii) make a major change to the building (e.g., building an extension), or (iii) change the use of the building. In determining whether to grant planning permission, the local authority has to assess whether the proposed development is consistent with (i) national policies, (ii) the local plan and (iii) the neighborhood plans (if any exists). Since most developments put a strain on existing infrastructure such as roads, schools and open spaces, the local planning authorities can impose a Community Infrastructure Levy (a charge which new developments pay, based on the size and type of development) to mitigate the impact of the proposed development. The Levy collected is then used to fund a wide range of infrastructure needed to support the development of the area. Alternatively, under Section 106 of the Town and Country Planning Act (1990), the local planning authority can also grant permission in return for some specified gain to the community. For example, the local planning authority may require that the developer provides a certain amount of affordable housing. This gives rise to a negotiation process between the local planning authority and the developer.

Applicants who have their plans rejected can appeal to the Secretary of State, via the Planning Inspectorate. The Planning Inspectorate then assigns one inspector to the case who decides whether to overturn the local planning authority's decision.[8]

In order to understand how inspectors are assigned to cases, we spoke to a representative from the Planning Inspectorate. The representative shared with us that the Planning Inspectorate uses an algorithm as well as human judgment to decide on the assignment of inspectors to cases. First, the algorithm uses the (i) complexity of the case (e.g., size of the development, whether the proposed development is in a "Greenbelt", agricultural area, "Area of Outstanding Natural Beauty", "Sites

[8] The Secretary of State has the power to take over the decision making from the Planning Inspectorate if the case raises particular issues that justify a Ministerial decision. In such cases, a planning inspector will submit a report and recommendation to the Secretary of State. Taking into account the inspector's assessment of the proposals, the Secretary of State will then make a decision.

of Special Scientific Interest", etc.) and (ii) how far away each inspector lives from the location of the proposed development to identify a list of suitable inspectors.[9] Second, a case worker then manually goes through the list and excludes inspectors who (i) already have a heavy case load, (ii) recently had a case in the area where the development is being proposed, (iii) live in the area of the proposed development, (iv) previously worked for a consultancy firm that is involved in the appeal, and/or (v) have personal difficulties such as illness. This means that conditional on these variables, the assignment of inspectors to cases is as good as random. This forms the basis of our empirical strategy to identify the causal effects of having a successful appeal.

3.3 Model

3.3.1 Overview

The aim of our model is to help interpret our results from the empirical analysis. We study developments in the UK that went through the appeals process. These projects are self-selected and would not be similar to projects that that did not appeal. Hence, we use the model to analyze this selection effect and how it changes with NIMBY-ism. The model is a sequential game between the developer (D), local planning authority (L) and inspector (N) played out in four stages:

Table 3.1: Stages of the game

Stage	Player	Action set	Consequence of action
0	Nature	Draws κ and η	All players observe κ, only N observes η
1	D	Propose a development plan s	Move to stage 2
2	L	Accept	s implemented
		Reject	Move to stage 3
3	N	Accept	s implemented
		Reject	No development
4	Housing market clears		

[9] The Planning Inspectorate maintains a list of inspectors which they classify based on the inspectors' suitability to assess cases according to the different levels of case complexity.

There is monopolist developer deciding how to develop a plot of land in a location. If developed, this land adds an exogenous amenity value of $\kappa \in (-\infty, \infty)$ to the location. The destruction of green spaces can be represented as a negative κ, while the removal of an abandoned building as a positive κ. A development plan is defined as the intensity of residential development (s), where $s \in [0, \infty]$. A higher s means more residents in the location which creates congestion. This is represented by an amenity value of $-\delta s$.[10] If the development is accepted, D's profits are $\pi = P_1(s, \kappa)s - \frac{1}{2}as^2$. $P_1(s, \kappa)$ is the market clearing price of the development and a captures the cost per unit intensity of a development.

We proceed to analyze the game backwards from stage 4:

3.3.2 Stage 4: Household behavior and market clearing

Prior to nature choosing κ: state of the location before new development

There is a mass of potential residents who are deciding between residing in the location or elsewhere (outside option):

$$\max_{\{\text{resident,outside option}\}} \quad C + \psi(x)$$

$$s.t. \quad P + C \leq W$$

W is wealth, C is composite consumption, and x and P are the stock of amenities and price of housing in the location. We assume $\psi'(x) > 0$ and $\psi''(x) < 0$. Potential residents have homogeneous utility from residing but heterogeneous outside utility. The mass of potential residents with an outside utility of u is $h(u)$ and we define $g(u) = \int_{\underline{u}}^{u} h(z)dz$. Market clearing requires that demand for housing equals supply ($\underline{s}$), or that there is a mass $\underline{s}$ of potential residents with utility of

[10]One way to think about congestion is that of more residents competing for scarce local services like schools and transport.

residing weakly better than their outside utility:

$$\underline{s} = g_0(u_0) \implies u_0 = g_0^{-1}(\underline{s})$$

Where u_0 is the outside utility of the marginal resident who is indifferent between residing or taking her outside option. u_0 also pins down the utility of residing for all residents, so we can rewrite the budget constraint to get the clearing price:

$$P_0 = W - C_0 = W - g_0^{-1}(\underline{s}) + \psi(x)$$

This says that housing prices are increasing in wealth and amenities but decreasing in supply of housing.

After development is built

After the development is built, θ proportion of residents are stuck and cannot move. We make this assumption to ensure that there are some residents that may be hurt by the development. If everyone can move away, then everyone will be better off after the development and the LPA will never reject developments.[11] In future versions of the paper we will allow residents to choose to leave by introducing wealth effects: movers will have to sell their properties at lower prices which reduces their wealth and hence, their outside utility. Now the problem for potential residents is:

$$\max_{\{\text{resident,outside option}\}} \quad C + \psi(x - \delta s + \kappa)$$

$$s.t. \quad P + C \leq W$$

[11]If all residents are free to move, or choose their outside options, market clearing implies $\underline{s} + s = g_0(u_1)$. Since $s \geq 0$ we must have $u_1 \geq u_0$, so the utility of all residents must be greater than before.

The only difference is that, after the development, the stock of amenities has changed by $-\delta s+\kappa$. Market clearing would imply

$$\underline{s}+(1-\theta)\underline{s}+s=(2-\theta)\underline{s}+s=g_1(u_1) \implies u_1=g_1^{-1}((2-\theta)\underline{s}+s)$$

where $g_1(u)$ is the cumulative mass of potential resident with outside options less than u, after excluding the stuck residents. Again, we can rewrite the budget constraint to get:

$$P_1(s,\kappa)=W-C_1=W-g_1^{-1}((2-\theta)\,\underline{s}+s)+\psi(x-\delta s+\kappa)$$

Notice that $P_{1s}(s,\kappa)<0$ because of the supply effect and disamenity effect.

3.3.3 Stage 3: Inspector decides whether to overturn the local planning authority's decision

The inspector (N) compares the social utility of accepting the project versus rejecting it. She observes η, which is the weight of current residents relative to the weight of new residents. So the social utility of accepting is:

$$\eta\theta\underline{s}[C_0+\psi(x-\delta s+\kappa)]+(s+(1-\theta)\underline{s})\,[C_1+\psi(x-\delta s+\kappa)]$$

Current residents, who are stuck, are weighted η, get their initial level of consumption (C_0) and enjoy the net stock of amenities. New residents, who are weighted 1, enjoy the same stock of amenities but get a higher level of consumption (C_1) because of the lower price of housing.

The social utility of rejecting is simply:

$$\eta\underline{s}[C_0+\psi(x)]$$

That is all residents stay and enjoy the initial level of utility.

D and L do not observe the value of η but know its distribution (cdf of $F(\eta)$). Given a value of s we can rewrite the condition for accepting as the range of values of η for which N will accept:[12]

$$\eta \leq \frac{(s + (1-\theta)\underline{s})\, g_1^{-1}((2-\theta)\,\underline{s} + s)}{\theta \underline{s}\,[\psi(x) - \psi(x - \delta s + \kappa)] + (1-\theta)\underline{s} g_0^{-1}(\underline{s})} \equiv I(s, \kappa)$$

Hence, the probability of a successful appeal is $F\left(I(s,\kappa)\right)$.

3.3.4 Stage 2: Local planning authority decides whether to accept or reject the developer's proposed plan

The local planning authority (L) knows that it can accept a development and it will be built. But if it rejects, N may still accept it. The social utility of L is similar to N but instead of η, she weighs current residents with a value of ω relative to new residents. She does not know the exact η but she knows the distribution of η. Assuming L is risk neutral, she will accept a development if and only if

$$\omega\theta\underline{s}[C_0 + \psi(x - \delta s + \kappa)] + (s + (1-\theta)\underline{s})\,[C_1 + \psi(x - \delta s + \kappa)] \geq \omega\underline{s}[C_0 + \psi(x)]$$

We can highlight some considerations of the developer using this inequality. When the development raises s there are three effects:

1. *Disamenity effect,* $\frac{\partial\psi(x-\delta s+\kappa)}{\partial s}$: lowers the value of amenities in the location and makes stuck residents worst off. This makes L less likely to accept.

2. *Weighting effect,* $\frac{\partial(s+(1-\theta)\bar{s})}{\partial s}$: raises the number of new residents who benefit from staying in the location and makes L more likely to accept.

3. *Price effect,* $\frac{\partial C_1}{\partial s}$: lowers the price of vacant and new housing, which raises the consumption and welfare of new residents. This makes L more likely to accept.

[12]Note that this expression is true if the denominator is positive. It must be positive if D sets s to be selected into appeals.

Which effect dominates will depend on the parameter values. For locations with a high δ, the disamenity effect should dominate, and a higher s would lower the social benefit. We can substitute the expressions from the market clearing equation to rewrite the acceptance condition:

$$(s + (1-\theta)\underline{s})\, g_1^{-1}((2-\theta)\,\underline{s} + s) \geq \omega\theta\underline{s}\,[\psi(x) - \psi(x - \delta s + \kappa)] + \omega(1-\theta)\underline{s}g_0^{-1}(\underline{s})$$

3.3.5 Stage 1: Developer proposes a development plan

Assuming that D is risk neutral, her expected utility from appealing is $F\left(I(s,\kappa)\right)\left(P_1(s,\kappa)s - \frac{1}{2}as^2\right)$. Notice that D can choose s to: (i) maximize profits subject to L accepting (honest); or (ii) maximize expected profits subject to L rejecting and D appealing (gaming). In our setup D always appeals after a rejection. This is reasonable because 83% of projects rejected by local authorities are appealed.[13]

Proposing an honest development

The honest developer's problem is:

$$\max_{s} \quad P_1(s,\kappa)s - \frac{1}{2}as^2$$

$$s.t.\ (s + (1-\theta)\underline{s})\, g_1^{-1}((2-\theta)\,\underline{s} + s) \geq \omega\theta\underline{s}\,[\psi(x) - \psi(x - \delta s + \kappa)] + \omega(1-\theta)\underline{s}g_0^{-1}(\underline{s}) \quad (IC)$$

$$P_1(s,\kappa)s - \frac{1}{2}as^2 \geq 0 \quad (PC)$$

$$P_1(s,\kappa) = W - C_1 = W - g_1^{-1}((2-\theta)\,\underline{s} + s) + \psi(x - \delta s + \kappa) \quad (MarketClearing)$$

The Incentive Constrain (IC) ensures L accepts the development, and the Participation Constraint (PC) ensures it is profitable to propose it. The IC may not bind if the unconstrained optimal s also

[13]This is estimated based on proposed major development from 4Q2012 to 4Q2017. Reasons for restricting our sample to major developments are in the data section.

satisfies the IC. Assuming that the PC is satisfied, the optimal profits can be written as

$$\pi^h\left(s^*,\kappa\right) = P_1(s^*,\kappa)s^* - \frac{a}{2}{s^*}^2$$

$$+\mu^*\left[\left(s^* + (1-\theta)\underline{s}\right)g_1^{-1}(\left(2-\theta\right)\underline{s} + s^*) - \omega\theta\underline{s}\left[\psi(x) - \psi(x-\delta s^*+\kappa)\right] - \omega(1-\theta)\underline{s}g_0^{-1}(\underline{s})\right]$$

where s^* is the optimally chosen development for the honest developer's problem.

Gaming the process

The gaming developer's problem is:

$$\max_{s} \quad F\left(I(s,\kappa)\right)\left(P_1(s,\kappa)s - \frac{1}{2}as^2\right)$$

$$s.t.\left(s + (1-\theta)\underline{s}\right)g_1^{-1}(\left(2-\theta\right)\underline{s} + s) \leq \omega\theta\underline{s}\left[\psi(x) - \psi(x-\delta s+\kappa)\right] + \omega(1-\theta)\underline{s}g_0^{-1}(\underline{s}) \quad (IC)$$

$$F\left(I(s,\kappa)\right)\left(P_1(s,\kappa)s - \frac{1}{2}as^2\right) \geq 0 \quad (PC)$$

$$P_1(s,\kappa) = W - C_1 = W - g_1^{-1}(\left(2-\theta\right)\underline{s} + s) + \psi(x-\delta s+\kappa) \quad (MarketClearing)$$

The IC in this case ensures L **rejects** the development. We have spoken to planning consultants in the UK and they do advise clients to propose developments, which will be rejected by the local authority. This tactic is employed when local planners are intransigent about compromising. Again the optimal profits can be written as

$$\pi^g\left(\tilde{s},\kappa\right) = F\left(I(\tilde{s},\kappa)\right)\left(P_1(\tilde{s},\kappa)s - \frac{1}{2}a\tilde{s}^2\right)$$

$$+\tilde{\mu}\left[\omega\theta\tilde{s}\left[\psi(x) - \psi(x-\delta\tilde{s}+\kappa)\right] + \omega(1-\theta)\underline{s}g_0^{-1}(\underline{s}) - \left(\tilde{s} + (1-\theta)\underline{s}\right)g_1^{-1}(\left(2-\theta\right)\underline{s} + \tilde{s})\right]$$

where $\tilde{s}$ is the optimally chosen development for the gaming developer's problem.

3.3.6 Selection into appeals

Developments that select into appeals are such that the profits from gaming are more than the profits from being honest, i.e., $\pi^g \geq \pi^h$. We hope to understand the range of κ that will result in appeals. To make some progress we make the following **assumptions**:

1. Define $g_0(u) = \alpha u$ and $g_0(0) = 0$. This assumes a demand elasticity of $-\alpha\frac{P_0(\underline{s})}{\underline{s}}$. A constant demand elascitiy is reasonable if the developments are not very large, or $s \ll \underline{s}$.

2. We consider two cases: (i) $\delta = 0$; and (ii) $\alpha \to \infty$. Case (i) is one where congestion is negligible, which would be the case if a location has excess capacity to absorb new residents. Case (ii) is one where demand is very elastic and so housing prices are not affected by supply. Case (i) shuts down the disamenity effect, and case (ii) shuts down the price effect. Having all three effects complicates the analysis by introducing discontinuous jumps in the constrained optimal profits. We will generalize these results in future versions of the paper.

Proposition 1 (Negative selection into appeals) *Given assumptions 1 and 2, in a subgame perfect Nash Equilibrium* $\exists\kappa^* : \kappa = \kappa^* \Rightarrow \pi^g = \pi^h$ *and* $\kappa < \kappa^* \Rightarrow \pi^g > \pi^h$ *and* $\kappa \geq \kappa^* \Rightarrow \pi^g \leq \pi^h$

Proof: See Appendix C.1.1. □

Appeals take place only when $\kappa < \kappa^*$. This says that the developments we observe in appeals are likely to be the ones with a more negative impact on amenities. For a honest development, large negative plot disamenities require intense residential developments to convince L to accept via the weighting and price effects. Given the convexity of costs, such developments quickly become unprofitable and the alternative of rolling a dice with a gaming development becomes more appealing.

Proposition 2 (Local power exacerbates selection into appeals) *In a subgame perfect Nash Equilibrium* $\omega' \geq \omega \Rightarrow \kappa^*(\omega') \geq \kappa^*(\omega)$

Proof: See Appendix C.1.2 □

An increase in ω makes the IC for the honest developer more binding because she has to give up more profits to placate the local planning authority. This increase, however, makes the IC for the gaming developer less binding because she wants the local planning authority to **not** accept. Therefore, local power encourages more developers to game. Intuitively, in locations with high levels of NIMBY-ism, even relatively benign projects end up being appealed. This is because developers have to give up too much of their profits to counter NIMBY-ism and so they are better off getting their plans rejected by the local authority and gambling on drawing an inspector who is less sympathetic towards locals' NIMBY behaviour.

3.3.7 Implications of selection on price regressions

Given propositions 1 and 2, we can now work out what our regressions are estimating. We consider two regressions: the first is an OLS regression $ln(p_i) = \beta_0 + \beta\mathbf{1}.\text{Successful Appeal}_i + e_i$, where the variation in successful appeals is generated by the random assignment of inspectors (or the random variable η). Here our coefficient of interest is β. The second is an 2SLS regression, where we run the regressions $ln(p_i) = \gamma_0 + \gamma_1\eta_i + u_i$ and $\mathbf{1}.\text{Successful Appeal}_i = \pi_0 + \pi_1\eta_i + v_i$, and our coefficient of interest is $\frac{\gamma_1}{\pi_1}$.

OLS versus IV regressions

The price change when we randomly assign η is: $\Delta P = E\left[P_1(\tilde{s}, \kappa) - P_0 | \kappa < \kappa^*\right]$.

Hence we know that β_1 in our OLS regression will be:

$$\frac{\Delta P}{P_0} = C \int_{\kappa_{pc}^g}^{\kappa^*} \left[\frac{-1}{\alpha}\left[\tilde{s} + (2-\theta)\underline{s}\right] + \left[\psi(x - \delta\tilde{s} + \kappa) - \psi(x)\right]\right] dG(\kappa)$$

where $C \equiv \frac{1}{G(\kappa^*) - G(\kappa_{pc}^g)} * \frac{1}{P_0}$, κ_{pc}^g is where the PC for the gaming problem binds and $G(\kappa)$ represent nature's distribution of κ. There are two main effects that influence the sign of this expression:

1. $\frac{-1}{\alpha}\left[\tilde{s}+(2-\theta)\underline{s}\right]$ is a supply effect, or a move along the demand curve, and is always negative.

2. $\psi(x-\delta\tilde{s}+\kappa)-\psi(x)$ is a demand shifter, and depends on whether $-\delta\tilde{s}+\kappa$ is greater or less than 0.

We can also work out that $\frac{\gamma_1}{\pi_1}$ in our IV regression will be:

$$C\int_{\kappa_{pc}^{g}}^{\kappa^*}\tilde{F}\Big(I(\tilde{s},\kappa)\Big)\left[\frac{-1}{\alpha}\left[\tilde{s}+(2-\theta)\underline{s}\right]+\left[\psi(x-\delta\tilde{s}+\kappa)-\psi(x)\right]\right]dG(\kappa)$$

where $\tilde{F}(I(\tilde{s},\kappa))=\frac{f(I(\tilde{s},\kappa))}{\int_{\kappa_{pc}^{g}}^{\kappa^*}f(I(\tilde{s},\kappa))dG(\kappa)}$. With no endogeneity issues, the difference between the OLS and IV estimates are the weights (1 versus $\tilde{F}(.)$). The OLS estimates the Average Treatment Effect (ATE), while the IV estimates a Local Average Treatment Effect (LATE), which skews towards developments that have a high increase in the probability of success from being assigned an inspector with a higher η. Under an assumption of homogeneity, that is κ is a constant, $\tilde{F}\Big(I(\tilde{s},\kappa)\Big)=1$ and the OLS and IV estimates are the same. This forms the basis of our regression analysis in Section 3.5.

3.4 Data

3.4.1 Description of datasets

The data we use are from four sources: (i) the Planning Inspectorate's Appeals Casework Portal (ACP); (ii) the UK Land Registry's Price Paid Database (PPD); (iii) the Domestic Energy Performance of Buildings Registers; and (iv) the Royal Mail Postal Address File (PAF).

Planning Inspectorate's Appeals Casework Portal (ACP). The ACP is the same database that the Planning Inspectorate caseworkers use to assign cases to inspectors and to manage cases. The ACP dataset contains decisions on appeals starting from October 1, 2012 to December, 28, 2018. It contains the following key variables:

Table 3.2: Summary of ACP dataset

Fields	Description
Local Planning Authority (LPA) details	Name of local planning authority (LPA), LPA code, LPA's case ID (the case ID allows us to search for documents related to appeal, e.g., site maps)
Key dates	Date when the appeal was received, started and decided
Inspector	Names of the inspectors
Development details	Residential or commercial, address, floor space, site area, number of residences whether the proposed development is in a "Greenbelt", "Area of Natural Beauty", "Sites of Special Scientific Interest", or agricultural area; whether it involves a conservation area or historical building
Others	Type of appeal, whether a planning consultant was hired, whether a bespoke timeline was agreed between the parties, whether the inspector ordered the local planning authority to foot the cost of the developer's appeal

Since the ACP data is also used by the Planning Inspectorate for casework management, most fields are accurate.[14] There is, however, one key field that is subjective because it is filled in by the developer – development type. Therefore, we do not use this variable to construct our sample.

UK Land Registry's Price Paid Database (PPD). For tax purposes, the PPD contains the registered price of all residential property transactions. The version of the dataset that we are using runs from January, 1, 2010 to January, 31, 2020. It includes the address of the property, the property type (i.e., detached, semi-detached, terrace or flat), whether the property is a new built and whether the property is freehold or leasehold. Crucially, the PPD does not include specific characteristics of the property such as the floor area or the number of rooms. Therefore, to obtain these variables, we use the Domestic Energy Performance of Buildings Registers.

Domestic Energy Performance of Buildings Registers. By law, Energy Performance Certificates (EPCs) are needed whenever a property is built, sold or rented. Besides the address of the property

[14] At the very least, they are consistent with the information the inspectorate had in deciding the appeal.

and its energy rating, this dataset also includes specific characteristics of the property like property type, floor area and number of rooms.

Royal Mail Postal Address File. This database contains all addresses and location of residential units in the UK. We use it to construct estimates of housing stock near the appeal site.

3.4.2 Sample for regression

We are interested in estimating the effect of a successful appeal on housing prices. To obtain the sample for our regression, we apply a number of sample restrictions.

Selection of appeals sample

Include only appeals involving residential developments that add 10 or more dwellings. There are a number of reasons why we apply this sample restriction. First, as noted in our background to land use regulations in the UK, almost any substantial changes one makes to a property requires planning permission – this includes, for instance, a loft conversion. This means that the ACP database includes appeals for a myriad of issues such as displaying advertisement on a building, house rear extensions and changes to commercial store front. Therefore, for this paper, we focus specifically on major residential developments, defined in the UK legislation as any development that adds 10 or more dwelling units.[15] We focus on housing because the debate on NIMBY-ism is generally about the restrictions in housing supply. Second, appeals involving smaller developments (less than 10) tend to comprise mainly of homeowners who are seeking to divide their house into two apartments to maximize rental yield or adding units to house family members. Such extensions and modifications are likely to be very different from a developer seeking to build residential units

[15]Town and Country Planning Order (2010) available at http://www.legislation.gov.uk/uksi/2010/2184/made. Note that this definition would include mixed use or primarily commercial developments that also add 10 or more residential units. In addition, there are certain requirements that "major developments" have to adhere to and these requirements make it easier for locals to exercise their power to influence the local planning authority. For example, the developer is required to give notice of the planned development at the development site, serve the notice to adjoining occupiers and publish the notice in a local newspaper. These requirements mean that locals are more likely to be aware of such developments, enabling them to voice their opinions and hold the members of the local planning authority accountable if they approve the development.

for sale. Third, there is a long tail of very small developments. For instance, more than 50 percent of residential developments add only one dwelling unit, and more than 75 percent of developments add two units. Since minor developments tend to have smaller externalities, we would be skewing our results towards finding zero effect if we include them in our sample.

Exclude variation in conditions and caravan parks. Variation in conditions are appeals that seek to amend the original plan approved by the local authority. For instance, a developer who planned for extensive landscaping, but now find it prohibitively costly, may appeal to the inspectorate to remove that part from the agreed plan. These appeals typically involve minor amendments and, if successful, barely change the development. We also exclude caravan parks because it is not clear whether they are residential in nature. Some may serve as long term lots for families living in caravans but others may serve as winter parking or tourist lots.

Choose earliest appeal if several are near each other. This reduces the need to account for cross-appeal effects in our regressions. We can also think of the number of subsequent appeals in the area as an outcome variable that is influenced by the earliest appeal. This may introduce endogeneity into our regressions.

Location plans available online. We had to geocode the planned development manually because many were developed on greenfield sites without existing addresses.

Include appeals for which inspector leniency is estimable. To construct the instrument for our IV regressions (see Section 3.5.3), we need a decent sample of cases per inspector to ensure estimates of inspector leniency converge to the true leniency. Thus we excluded appeals for which the inspector did not have many cases. The sample counts following the various restrictions are summarized here and Table 3.3 shows the summary statistics of the appeals data after applying these restrictions.

1. Major residential appeals 2012-2018 (i.e., $\geq$ 10 dwellings) $\rightarrow$ **4,324 appeals**

2. Exclude variation in conditions, caravan parks, choose earliest appeal → **4,211 appeals**

3. Able to manually geocode → **3,513 appeals**

4. Can estimate inspector leniency → **3,121 appeals**

Table 3.3: Summary statistics (appeals data)

	Mean	SD	Min	Max	N
Year appeal started	2015.42	1.65	2012	2018	3121
Year appeal decided	2015.86	1.61	2012	2018	3121
No. of dwellings	56.19	100.93	10	4022	3121
Site area (hectres)	6.88	141.55	0	6780	2987
East Midlands	0.11	0.31	0	1	3121
East of England	0.14	0.35	0	1	3121
London	0.09	0.28	0	1	3121
North East	0.03	0.16	0	1	3121
North West	0.09	0.29	0	1	3121
South East	0.24	0.43	0	1	3121
South West	0.15	0.36	0	1	3121
West Midlands	0.09	0.29	0	1	3121
Yorkshire and The Humber	0.05	0.23	0	1	3121
Appeal successful	0.43	0.50	0	1	3121

Selection of price paid sample

Include only resale transactions of residential properties that are ever within 1km of any appeal. The reason for including only resale transactions is because we want to see how the appeals affect existing properties and not the new properties that are being built. As for the 1km restriction, we apply this because if we were to increase the radius beyond 1km, we end up having many properties that are linked to multiple appeals. In addition, the effect of having an appeal near-by is likely to decay with distance. If we define too big a radius, we would be skewing our results towards finding zero effect. In future work, we will be checking if our results are robust to using different distance thresholds.

Exclude "others" category of residential transactions. The dataset includes residential-related transactions such as garages and parking spaces that we exclude.

Include only resale transactions that took place within 3 years before or after an appeal. Since the appeals dataset only starts on October 1, 2012, we have fewer and fewer resale transactions beyond 3 years of an appeal. Statistical inference might become unreliable if the number of resale transactions in certain years before or after an appeal becomes too small. We will show as a robustness check that our results are stable to the choice of year bandwidths.

Trimmed top and bottom 1% in prices. Regressions may be sensitive to outliers and we trimmed extreme prices to reduce that sensitivity. Our results hold with an untrimmed or a Winsorised sample.

Table 3.4 shows us the summary statistics of the resale transactions data.

Table 3.4: Summary statistics (resale transactions data)

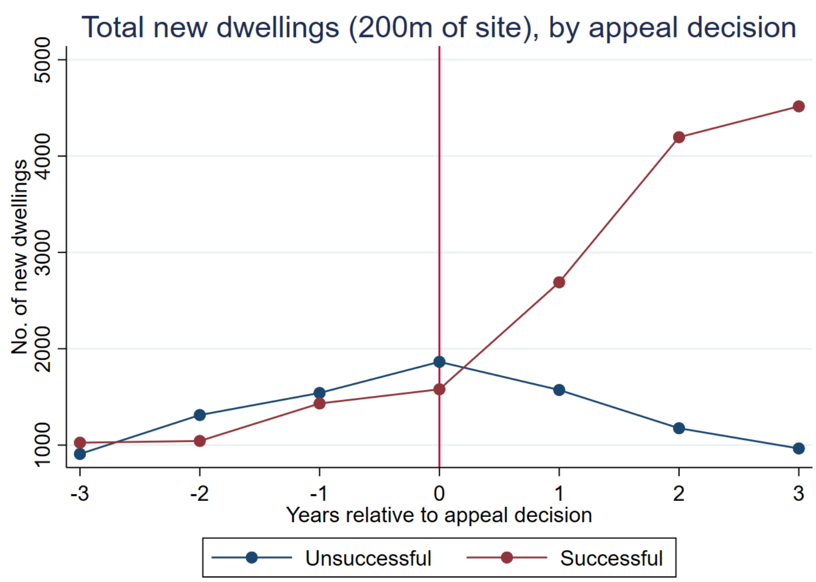
	Mean	SD	Min	Max	N
Year of transaction	2015.88	2.07	2010	2020	1238407
Transacted price	297463.88	209187.77	54950	1504000	1238407
Terrace	0.29	0.45	0	1	1238407
Flat	0.27	0.45	0	1	1238407
Semi-detached	0.24	0.42	0	1	1238407
Detached	0.20	0.40	0	1	1238407
Freehold	0.70	0.46	0	1	1238407
Distance from first appeal	631.36	246.15	1	1000	1238407

3.4.3 What happens after a successful appeal?

So far we have assumed that successful appeals lead to new residential developments at the appeal site. Concretely, what it does lead to is an option to develop the appeal site. Developers may choose not to develop at all if property markets are poor.[16] Figure 3.1 plots the number of new dwellings, within 200m of the center of appeal sites, before and after an appeal decision, and indicates that options do translate to new housing. Controlling for location and year fixed-effects, this translates to a 0.42% increase in housing supply near the appeal site.

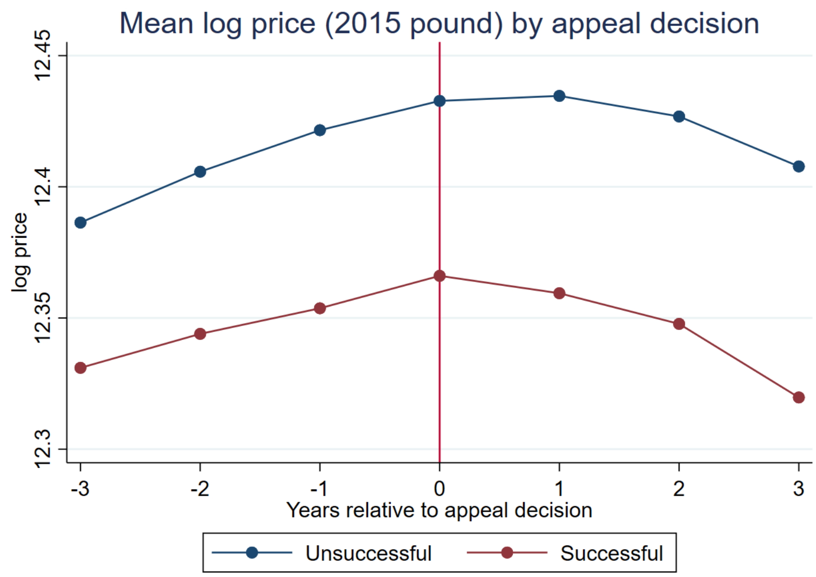
Figure 3.2 plots the average log real price by appeal decision. We see that successful appeals tend to be in neighborhoods with lower house prices. This suggests neighborhood characteristics, which determine market prices, may affect appeal decisions. For instance, properties near a green belt are priced higher, due to the supply restrictions; and the bar for an appeal there would also be higher. However, we see that the price trends prior to appeal decisions are similar across both groups.

Figure 3.1: Comparison of new dwellings in appeal sites, by success of appeal



[16]What they can do, however, is to sell the option to other developers. Regardless of market conditions, developers who have gone through the appeals process would have sunk in a lot of money (e.g., purchasing the land, drafting plans, hiring architects and consultants) and would usually develop the site except in exceptional circumstances.

Figure 3.2: Unconditional mean log price by appeal decision



3.5 Empirical strategy

3.5.1 Overview

To estimate the causal effect of overturning the local authority's decision on housing prices, we can run the following DiD regression:

$$ln(price_{ijt}) = \alpha_j + \delta_t + \beta Success_i \times Postappeal_t + \gamma' X_{ijt} + \epsilon_{ijt} \quad (3.1)$$

$ln\,(price_{ijt})$ is the log price of property i within 1km of appeal j in year-month t. $Success_i$ is an indicator variable that denotes whether property i is **ever** within 1km of a **successful** appeal. $Postappeal_t$ is an indicator variable for the period that the appeal takes place and periods after the appeal. X_{ijt} is a vector of individual property characteristics such as property type, whether the property is leasehold or freehold, the number of rooms, floor area, the property's current and potential energy efficiency, whether the property was built before 2012 and the supply trend in the 1km vicinity of the appeal prior to the appeal decision. α_j are fixed-effects for properties around appeal j. Finally, δ_t are time fixed effects. These consists of calendar year-month fixed-effects and relative year fixed-effects. For example, the relative year fixed-effect takes on the value of 1 if the sale of the property is within 1 year after the appeal. We cluster the standard errors at the outward code level. Using the example of the postcode SE16 7BB, Table 3.5 shows how the varying levels of postcode granularity are defined in the UK. The outward code thus corresponds to a district in the UK (roughly the size of a town or part of a large town).

Table 3.5: Postcode format in the UK

Postcode			
Outward code		Inward code	
Area	District	Sector	Unit
SE	16	7	BB

3.5.2 Limitations to difference-in-differences

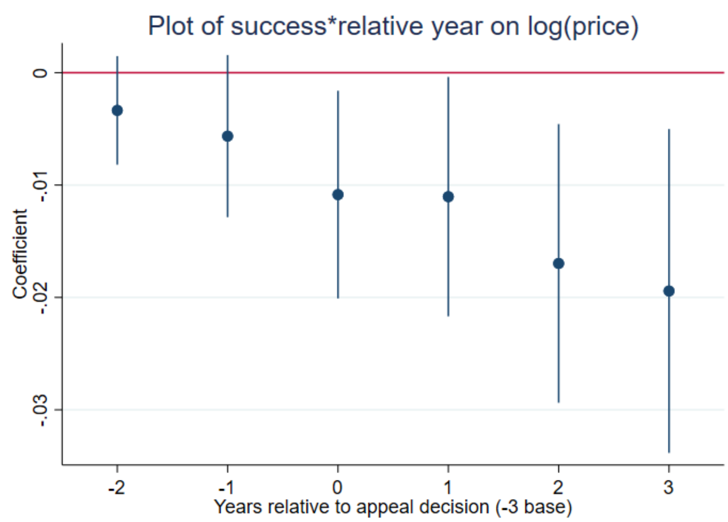
Figure 3.3 graphs the coefficients (with the 95% confidence intervals) from a dynamic DiD regression (with controls) that compares the prices, relative to three years before the appeal decisions, between properties near successful versus unsuccessful appeal sites.[17] There is no significant difference in price trends between the two groups, suggesting that a DiD specification may give a consistent estimate of the Average Treatment Effect (ATE) of having a successful appeal development nearby. The values of the coefficients used to plot Figure 3.3 are reported in Appendix Table C1.

However, even if equation 3.1 exhibits parallel pre-trends, we might still worry about unobserved time-varying neighborhood characteristics that vary **after** the appeal and also affect housing prices. For example, after an appeal (regardless of whether it was successful or not), residents in high NIMBY areas might take the opportunity to quickly come up with a neighborhood plan so as to

[17] This regression includes the standard controls we use in our main specifications, as outlined in equation 3.1.

make it difficult for other developers to develop new housing in their neighborhood in the future. Therefore, to address such potential selection bias that happens after an appeal, we adopt an IV approach. We exploit the fact that appeals are randomly assigned to inspectors (conditional on the assignment criteria used by the Planning Inspectorate which we elaborated on in Section 3.2). In addition, some inspectors are systematically more lenient that others. Taken together, these lead to random variations in the probability that an appeal will be successful based on which inspector the appeal is assigned to. Since there is no evidence of pre-trends in our DiD regression, when reporting our empirical results, we report both the DiD and IV results.

Figure 3.3: Difference in prices relative to 3 years before appeal decision



3.5.3 Instrumental variable calculation

We measure the average leniency of an inspector based on the appeal success rate for all the other randomly assigned cases that the inspector handled.[18] These cases include both past and future appeals but not the existing appeal. This leave-out measure is important because it avoids introducing mechanical reverse causality. To construct the instrument, we follow the existing

[18]Although our regression sample involves only major dwellings, we also use the success rate for appeals involving non-major dwellings to construct the instrument. The purpose of doing this is to improve the power of the instrument.

literature on expert leniency and regress whether the appeal is successful on the variables which we know that the Planning Inspectorate uses to assign the inspectors to cases. The residual from this leave-one out regression is our leniency measure. The assignment criteria variables include the workload of the inspectors, characteristics of the area around the appeal site and the complexity of the case.[19] Controlling for the assignment criteria is important because it accounts for the fact that randomization by the Planning Inspectorate occurs within the pool of available and suitable inspectors.

While we use all the available data to estimate the leniency of the inspectors, after applying our sample restrictions, for our regression sample, we have 412 inspectors who were assigned to appeals involving major dwellings. Each of these inspectors presided over an average of around 65 appeals. The highest number of appeals presided over by an inspector is 313 and the smallest number of appeals presided over by an inspector is 1. Essentially, we estimate the leniency of inspectors using her mean residualized probability of approving an appeal. This estimate would be very noisy for inspectors with only a few appeals. For instance, the leniency would be estimated off 1 appeal if an inspector only judged 2 appeals in total.[20] Therefore, we did not use inspectors, in our regressions, who presided over <30 appeals. This excludes 31.6% of inspectors from our sample.[21]

Figure 3.4 shows the identifying variation in our data. Controlling for the vector of assignment variables used by the Planning Inspectorate, the inspector leniency measure ranges from -0.25 to 0.30 with a standard deviation of 0.09. The histogram suggests that there is a wide variation in whether an inspector is likely to allow an appeal to be successful.

---

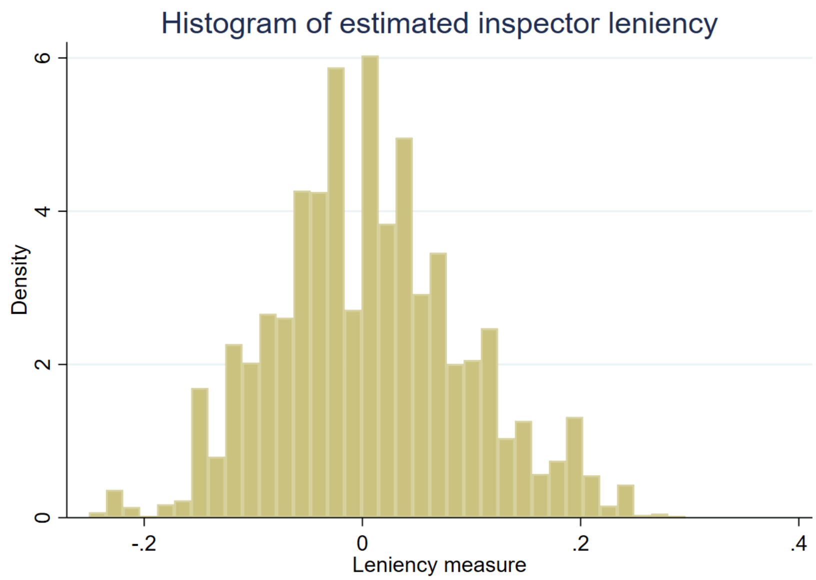
[19]To proxy for the complexity of the case, we use variables such as the size of the development that is being appealed, whether the appeal relates to other cases, whether a bespoke timeline was agreed between the parties, the total number of days that it took for the inspector to come to a decision, and key words used to describe the appeal (e.g., demolition, commercial, facilities).

[20]Indeed, the predictive power of leniency on approval, without dropping inspectors, is very poor.

[21]Our results are robust to using various arbitrary cut-offs.

Finally, we interact the leniency measure with the post appeal dummy. $Leniency_i \times Postappeal_t$ is thus our instrument for $Success_i \times Postappeal_t$.

Figure 3.4: Distribution of inspector leniency (regression sample)



3.5.4 Validating the IV strategy

Relevance of Instrument

We estimate the following linear probability model to examine the first-stage relationship between inspector leniency and whether an appeal is successful:

$$Success_i \times Postappeal_t = \alpha_j + \delta_t + \rho Leniency_i \times Postappeal_t + \theta' X_{ijt} + \upsilon_{ijt} \tag{3.2}$$

where X_{ijt} is a vector of control variables that is the same as in equation 3.1.

Table 3.6 presents the first-stage results. The estimates are highly significant, suggesting that being assigned to an inspector who is 10 percentage points more lenient increases the probability of the appeal being successful by around 6.6 percentage points. Furthermore, the first-stage has a KP F-statistic value of 24, suggesting that we can reject the null hypothesis of weak instruments.[22]

Table 3.6: First-stage estimates

	(1)	(2)
VARIABLES	Success*Postappeal	
Leniency*Postappeal	0.664***	0.664***
	(0.135)	(0.135)
Observations	1,176,342	1,176,164
Year-by-Month FE	✓	✓
Appeal FE	✓	
Appeal X Postcode Sector FE		✓
Controls	✓	✓
KP F-stat	24.04	24.07

Notes: Standard errors clustered at outward code level.
Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Validity of Instrument

Conditional Independence. In order for inspector leniency to be a valid instrument, the assignment of appeals to inspectors must not be correlated with variables which also affect the outcome variable (housing prices). We aggregate the data by properties for all the past transactions that are within 1km of **future** appeals and run a balancing test.

Table 3.7 shows that past transaction characteristics of properties within 1km of an appeal are not predictive of inspector leniency. All of the estimates are close to zero and all of them are statistically insignificant at the 5% level. More importantly, the variables are not jointly significant with a p-value of 0.246. This provides empirical support that conditional on the Planning Inspectorate's assignment rules, inspectors are randomly assigned to cases. The random assignment of inspectors

[22]The results presented in Table 3.6 include all controls. In Table C2, we show how the coefficient estimate changes when we move from a specification with just the basic fixed effects to one where controls are added.

(conditional on the Planning Inspectorate's assignment criteria) gives us consistent estimates of the reduced form effect of inspector leniency on housing prices. However, interpreting the IV estimates as the causal effects of a successful appeal on housing prices requires two further assumptions.

Table 3.7: Balancing test

	(1)		
VARIABLES	Inspector Leniency	VARIABLES	Inspector Leniency
ln(price)	0.002	Potential energy D	-0.003
	(0.003)		(0.002)
ln(floor area)	-0.001	Potential energy E	-0.003
	(0.003)		(0.003)
3 rooms	0.000	Potential energy F	-0.005
	(0.001)		(0.003)
4 rooms	-0.000	Potential energy G	-0.002
	(0.002)		(0.004)
5 rooms	0.001	Detached house	-0.000
	(0.002)		(0.003)
6 rooms or more	-0.000	Semi-detached house	-0.002
	(0.002)		(0.002)
Current energy B	-0.025	Flat	0.009
	(0.025)		(0.005)
Current energy C	-0.029	Leasehold property	-0.007
	(0.025)		(0.005)
Current energy D	-0.029	Built before 2012	-0.002
	(0.025)		(0.004)
Current energy E	-0.028	Neighbourhood density trend	-0.000
	(0.025)		(0.000)
Current energy F	-0.027		
	(0.025)		
Current energy G	-0.027	Observations	594,211
	(0.025)	Adjusted R-squared	0.004
Potential energy B	-0.004*	Time FE	✓
	(0.002)		
Potential energy C	-0.005*	F-stat for joint test	1.186
	(0.002)	p-value for joint test	0.246

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Exclusion. This restriction requires that inspector leniency affects housing prices (outcome variable) only through its effect on whether an appeal is successful. Although there is no direct way to test the exclusion restriction, the fact that inspectors are randomly assigned to cases lends support to the exclusion restriction.

Monotonicity. If the causal effect of a successful appeal is constant across all cases, then the instrument only needs to satisfy the exclusion assumption. However, with heterogeneous effects, monotonicity must also be assumed. Monotonicity gives the IV estimate a local average treatment effect interpretation – the average causal effect among the subgroup of cases that would have received a different appeal decision had the case been assigned to a different inspector. In our setting, the monotonicity assumption requires that appeals that are ruled successful by a strict inspector would also be ruled successful by a lenient inspector.

One testable implication of the monotonicity assumption is that the first-stage estimates should be non-negative for any subsample. To test this, we split the sample into subsamples based on (i) geographic regions; (ii) year of appeal decision; and (iii) number of dwellings in the development. Table 3.8 shows that for all of these subsamples bar one, the first-stage estimates are positive. For the region, West Midlands (region 11 in Table 3.8), which accounts for 9% of appeals, the coefficient is negative but not statistically significant. Excluding this region does not change our main results, so we do not think this is strong evidence that the monotonicity assumption is violated.

Table 3.8: Testing monotonicity assumption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES				Success*Postappeal			
Leniency*Postappeal	1.098***	0.390	1.147***	1.448**	0.966**	0.399	0.804**
	(0.418)	(0.416)	(0.291)	(0.673)	(0.492)	(0.273)	(0.340)
Observations	102,782	136,421	236,083	21,726	99,084	268,379	156,215
KP F-stat	6.878	0.878	15.49	4.635	3.856	2.135	5.583
Region	1	2	3	4	5	8	9

	(8)	(9)	(10)	(11)	(12)	(13)
VARIABLES			Success*Postappeal			
Leniency*Postappeal	-0.660	0.528	1.049***	0.368*	1.001***	0.379**
	(0.446)	(0.447)	(0.191)	(0.189)	(0.201)	(0.184)
Observations	81,220	74,007	474,035	702,307	578,132	598,210
KP F-stat	2.188	1.394	30.14	3.793	24.75	4.250
Region	11	12				
Decision year			2012-2015	2016-2018		
No. of dwellings					10-25	>25

Notes: All regressions include time and appeal fixed effects as well as controls. Standard errors clustered at outward code level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1

3.6 Results

Before presenting the DiD (OLS) and IV estimates formally, we show in Figure 3.5 top panel, the conditional mean of the probability of a successful appeal by inspector leniency, and in the bottom panel the conditional mean log price by inspector leniency. This allows us to visually inspect the underlying variation in our data. The top panel is akin to a dynamic first-stage regression with controls. The only difference is that for presentation purposes, we plot inspector leniency in terms of being above or below the median. Similarly, the bottom panel is akin to a dynamic reduced-form regression with controls. The top panel reveals that inspector leniency has a high predictive power as to whether an appeal will be successful. The bottom panel shows us that prices of resale transactions do indeed diverge based on whether a lenient or strict inspector is assigned to a nearby appeal.

Figure 3.5: Conditional dynamic first-stage and reduced form by inspector leniency

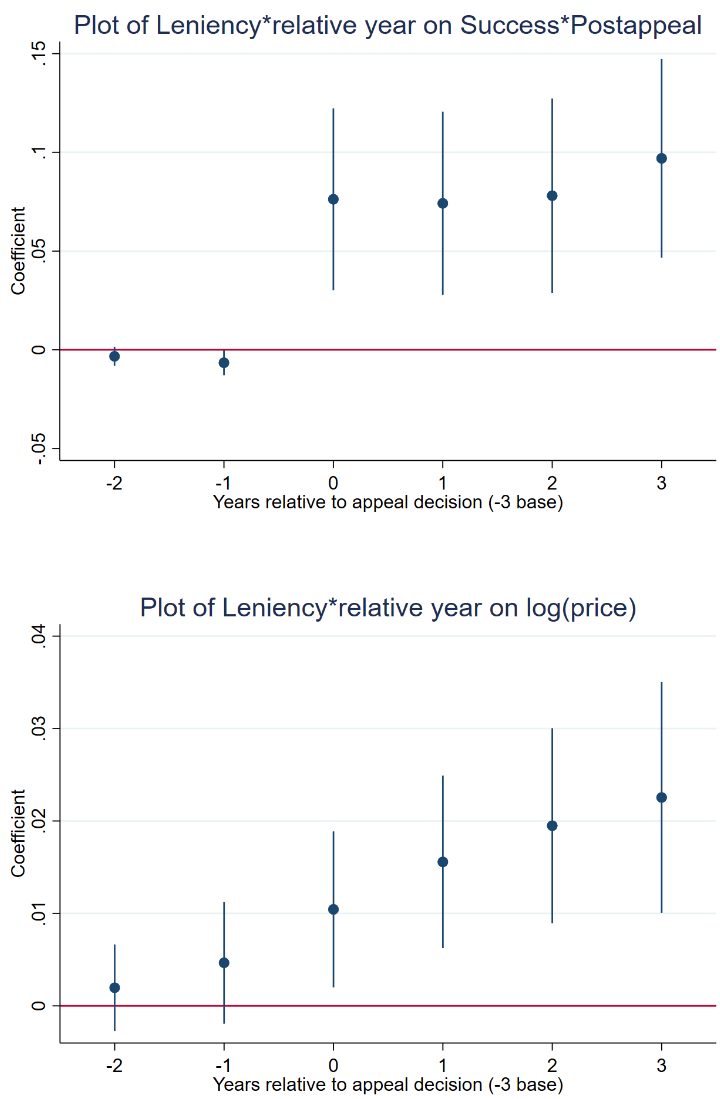


Table 3.9 formally presents the DiD (OLS) and IV estimates of the effect of overturning the local authority's decision on housing prices. Columns 1 to 2 report OLS estimates with all the controls. In the latter columns we also include appeal-by-postcode sector FE instead of appeals FE. The OLS estimates suggest that a successful appeal has a small negative effect of around 1% on housing prices (compared to an unsuccessful appeal). The 95% confidence interval is able to rule

out an impact more negative than -1.8%. The rejection rate, of major developments, for all local authorities in the UK was about 20% from 2012 to 2017. This suggests that local authorities were rejecting only the most egregious developments. Hence, it is surprising that we can rule out a large negative impact on prices in the neighborhood.

The IV estimates are presented in columns 3 and 4. They suggest that, instead of depressing prices, overturning the local authorities' decision actually increased the value of properties in the neighborhood by around 6%. The coefficient is significant at the 10% level, and its 95% confidence interval is able to reject an impact more negative than -0.5%. Similar to the OLS, we are not finding that developments that are rejected have a large detrimental impact on the neighborhood.[23]

Table 3.9: Effect of overturning the Local Authority's decision

	(1)	(2)	(3)	(4)
	OLS		IV	
VARIABLES	ln(price)			
Success*Postappeal	-0.010**	-0.010**	0.060*	0.062*
	(0.004)	(0.004)	(0.033)	(0.033)
Observations	1,177,861	1,177,683	1,176,342	1,176,164
R-squared	0.843	0.856	0.623	0.623
Time FE	✓	✓	✓	✓
Appeal FE	✓		✓	
Appeal X Postcode sector FE		✓		✓
Controls	✓	✓	✓	✓
KP F-stat			24.04	24.07

Notes: Standard errors clustered at outward code level.
Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

The 95% confidence intervals of the OLS and IV estimates overlap, so we cannot reject the null that the IV and OLS are different. This is mostly because the IV estimates are not estimated very

[23]The results presented in Table 3.9 include all controls. In Tables C3 (OLS) and C4 (IV), we show how the coefficient estimate changes when we move from a specification with just the basic fixed effects to one where controls are added.

precisely. However, the direction of the estimates is different and the magnitude of the IV estimate is a substantial positive number. We think this is due to the IV estimating a LATE instead of the ATE. The OLS estimates tell us that the ATE of overturning local authorities' decision is a small negative impact on the neighborhood. But for a subsample of these projects, those that are quite marginal and hence are very susceptible to the leniency of inspectors, the impact may be positive.

We interpret these results as suggestive of NIMBY-ism. First, the most egregious development in the UK have, at most, a small negative impact if allowed. Second, a subsample of these projects may actually benefit the local neighborhood, which suggests NIMBY-ers may be blocking any project that would change their neighborhood.

To further support this interpretation, we look at comments lodged by residents close to the projects. We randomly sample 68 projects in our sample (rejected projects) from London. For each project, we randomly select a project near to it (within 1 to 2km) as a comparison that was accepted by the local authority.[24] The mean number of comments lodged for rejected projects is 3.07 as compared to 0.72 for accepted projects (p-value of 0.0375). To the extent that comments are used to discourage projects, this is evidence for NIMBY-ism.

3.6.1 Heterogeneous effects

We attempt to decompose the impact by distance from the appeal site. Columns 1 and 3 of Table 3.10 show the estimated impact when we restrict the sample to properties within 500m of the appeal site, while Columns 2 and 4 of Table 3.10 show the estimated impact when we restrict the sample to properties 500m to 1km from the appeal site. Both the OLS and IV results suggest little heterogeneity in the impact by distance from the appeal site.[25]

[24]We did not look at projects within 1km of an appeal site because it may be affected by the appeal.

[25]The results presented in Table 3.10 include all controls. In Tables C5 (500m sample) and C6 (500m to 1km sample), we show how the coefficient estimate changes when we move from a specification with just the basic fixed effects to one where controls are added.

Table 3.10: Effect by distance from appeal site

	(1)	(2)	(3)	(4)
	OLS		IV	
VARIABLES	ln(price)			
Success*Postappeal	-0.009**	-0.010**	0.051	0.073*
	(0.005)	(0.004)	(0.031)	(0.038)
Observations	360,933	816,622	360,563	815,474
R-squared	0.857	0.861	0.615	0.619
Time FE	✓	✓	✓	✓
Appeal X Postcode sector FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Sample	<500m	500m to 1km	<500m	500m to 1km
KP F-stat			28.96	19.85

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 3.11 explores heterogeneity by the supply shock from a successful appeal. We define the supply shock as the number of new dwellings that would be introduced by the appeal, divided by the total number of dwellings within 1km of the appeal site, in the year prior to the appeal decision. Small and large shocks are categorized based on whether they were below or above the median magnitude of supply shocks. For the OLS, there was no significant heterogeneity by the size of the supply shocks. For the IV, small supply shocks have a greater positive impact on prices in the neighborhood than larger supply shocks. This is consistent with a simple demand and supply model, as a smaller supply shock suggests a small movement along the demand curve. However, a positive impact also suggests a shift in the demand curve. This could be because the new developments also bring in new or better amenities to the neighborhood. In the case of developments with a small supply shock, the demand shifters overwhelmed the supply shock; while for larger developments the two seem to cancel out. This interpretation is for the IV and hence applies only to marginal developments (LATE). Also, the 95% confidence intervals of the two samples overlap, so we cannot reject the null of no differences.[26]

Table 3.11: Effect by supply shock

	(1)	(2)	(3)	(4)
	OLS		IV	
VARIABLES	ln(price)			
Success*Postappeal	-0.009*	-0.007	0.070**	-0.011
	(0.005)	(0.005)	(0.035)	(0.083)
Observations	913,192	264,491	912,619	263,545
R-squared	0.860	0.831	0.618	0.646
Time FE	✓	✓	✓	✓
Appeal X Postcode sector FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Sample	Small shock	Large shock	Small shock	Large shock
KP F-stat			24.70	1.838

Notes: Standard errors clustered at outward code level.
Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1

3.6.2 Mechanisms

One limitation of the data is that we only observe transaction prices. However, the outcome of an appeal may affect the choice to buy or sell a nearby property (selection into transaction). For instance, richer households in the neighborhood, who have the finances to move quickly, may choose to sell after a successful appeal as they anticipate that the new supply would depress the value of their housing asset. If these properties are also more valuable (e.g., bigger and more luxurious), then our IV result of a price increase from a successful appeal might be due to this selection into transaction.[27] We check if this is an issue by regressing some characteristics of the

[26] The results presented in Table 3.11 include all controls. In Tables C7 (small shock) and C8 (large shock), we show how the coefficient estimate changes when we move from a specification with just the basic fixed effects to one where controls are added.

[27] We do control for factors related to the characteristics of transacted properties in all our regressions. However, if characteristics are endogenous to appeal decisions then they should be outcome variables instead of controls (bad controls problem).

transacted properties on the treatment variable ($Success_i * Postappeal_t$). If there is selection into transaction, then we should expect the types of properties being transacted to be different between neighborhoods with successful and unsuccessful appeals.

Table 3.12 summarizes our findings on three property characteristics: (i) floor area; (ii) number of rooms; and (iii) energy rating (a higher rating is more energy efficient). The first two measures size and the third is a proxy for the type of materials used in the property.[28] Both the OLS and the IV regressions cannot reject the null that there are no differences in transacted properties near successful versus unsuccessful appeals. We interpret this as evidence that selection into transaction is not a major driver of our main results above.

Table 3.12: Possible mechanisms

	(1)	(2)	(3)	(4)	(5)	(6)
		OLS			IV	
VARIABLES	ln(Area)	No. rooms	Energy rating	ln(Area)	No. rooms	Energy rating
Success*Postappeal	0.002	0.004	0.005	0.012	-0.032	0.029
	(0.001)	(0.007)	(0.004)	(0.013)	(0.059)	(0.032)
Observations	1,213,468	1,177,683	1,177,683	1,211,946	1,176,164	1,176,164
R-squared	0.417	0.395	0.479	0.287	0.261	0.400
Time FE	✓	✓	✓	✓	✓	✓
Appeal X Post sect. FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
KP F-stat				24.19	24.07	24.07

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Is there evidence of demand shifters? To answer this, we collect annual point-of-interest (POI) data by location and test if successful appeals led to higher counts of these points. We categorize

[28]This is an imperfect measure of "quality" because there may be luxurious building materials that are not energy efficient, and rich households may be willing to pay higher heating costs to maintain them. We recognize this and are simply interested to test if the materials used are different as opposed to better.

POIs into 3 categories, summarized in Table 3.13. These categories reflect the possible amenities that major developments could introduce: (i) integrated developments often include commercial spaces for local businesses; (ii) population growth in an area may attract more businesses; and (iii) developers may be asked to contribute to local infrastructure by building them or via a tax.

Table 3.13: POI Categories

Category	POIs included
Local services	Accommodation, Eating and Drinking, Retail
Local economy	Commercial services, Manufacturing and Production
Local infrastructure	Education and Health, Public Infrastructure, Transport

We use the count of POIs in each category as an outcome variable and study if there were more POIs within 1km of a successful appeal site (Table 3.14). Our results are noisy and we cannot make strong conclusions.[29] However, there is suggestive evidence that the count of local services increased after successful appeals (IV regression significant at 10% level).

Table 3.14: Possible demand shifters

	(1)	(2)	(3)	(4)	(5)	(6)
		OLS			IV	
VARIABLES	Local serv.	Local econ.	Local infra.	Local serv.	Local econ.	Local infra.
Success*Postappeal	-0.399	1.537	0.220	11.419*	6.340	-0.639
	(0.643)	(1.104)	(0.871)	(6.743)	(7.113)	(9.057)
Observations	18,654	18,654	18,654	18,624	18,624	18,624
R-squared	0.995	0.994	0.991	-0.046	-0.002	-0.000
Time FE	✓	✓	✓	✓	✓	✓
Appeal FE	✓	✓	✓	✓	✓	✓
KP F-stat				27.76	27.76	27.76

Notes: Standard errors clustered at outward code level.
Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

If we disaggregate local services (Table 3.15), we see that the increase is coming from retail shops. This is evidence that the positive impact from marginal appeals (LATE), may be driven by

[29] This is due to the high annual turnover in POIs: businesses open and shut down, bus stops are added etc.

an improvement in local amenities which increased demand for the neighborhood.

Table 3.15: Demand shifters within local services

	(1)	(2)
	IV	
VARIABLES	Accommodation, Eating and Drinking	Retail
Success*Postappeal	1.162	10.258*
	(2.594)	(5.605)
Observations	18,624	18,624
R-squared	-0.001	-0.065
Time FE	✓	✓
Appeal FE	✓	✓
KP F-stat	27.76	27.76

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Robustness Checks

Next, we subject our results to a series of robustness checks. First, we show that our results are robust to controlling for location specific linear year trends as well as clustering our standard errors at the appeal and distance level. Table 3.16 shows the results for three levels of clustering: (i) appeal-by-postcode sector neighborhood; (ii) postcode sector; (i) 1km grids. The 1km grids are arbitrarily created grids of 1km by 1km across the whole of the UK. Inference on the OLS does not change much across these different clusters. The standard errors for the IV are noisier under the appeals neighborhood cluster (column 4), and we can no longer reject a null of zero effect at the 10% level. However, the 95% confidence interval still allows us to reject an impact more negative than -1.7%. Therefore, our original inference of no big negative impact still stands. Other methods of clustering help improve precision and allow us to reject a null of zero impact.

Table 3.16: Robustness to different clustering

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS			IV		
VARIABLES	ln(price)					
Success*Postappeal	-0.010**	-0.010***	-0.010***	0.062	0.062**	0.062***
	(0.004)	(0.003)	(0.003)	(0.039)	(0.025)	(0.022)
Observations	1,177,683	1,177,683	1,177,683	1,176,164	1,176,164	1,176,164
R-squared	0.856	0.856	0.856	0.623	0.623	0.623
Time FE	✓	✓	✓	✓	✓	✓
Appeal X Post sect. FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
Clusters	Appeal-Post sect.	Post sect.	1km grids	Appeal-Post sect.	Post sect.	1km grids
KP F-stat				18.65	42.77	60.39

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

In Table 3.17 we check whether our results are robust to different appeal neighborhood and appeal-by-postcode sector specific time trends. The trends we include are estimated off a similar variation as our treatment variable $Success_i * Postappeal_t$ and we should hence expect our coefficients of interests to attenuate towards zero. This is indeed what happens and none of the estimates can reject a null of zero impact. However, the 95% confidence intervals allow us to reject a null of an impact more negative than -1% (OLS) and -1.9% (IV).

Finally, we trim off the top and bottom values in our instrument (inspector leniency), to check if our IV results are sensitive to outlier values (Table 3.18). After trimming 1% of extreme values (column 1), the IV estimates remain positive, at 7.2%, and are now significantly different from zero at the 5% level. Trimming 5% of extreme values causes the estimate to remain positive but attenuate towards zero.

Table 3.17: Robustness to different trends

	(1)	(2)	(3)	(4)
	OLS		IV	
VARIABLES	ln(price)			
Success*Postappeal	-0.004	-0.004	0.026	0.028
	(0.003)	(0.003)	(0.023)	(0.023)
Observations	1,177,861	1,177,683	1,176,342	1,176,164
R-squared	0.846	0.860	0.629	0.631
Time FE	✓	✓	✓	✓
Appeal FE	✓		✓	
Appeal trends	✓		✓	
Appeal X Postcode sector FE		✓		✓
Appeal X Postcode sector trends		✓		✓
Controls	✓	✓	✓	✓
KP F-stat			26.66	26.58

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 3.18: IV results after trimming

	(1)	(2)
VARIABLES	ln(price)	
Success*Postappeal	0.072**	0.040
	(0.033)	(0.034)
Observations	1,155,254	1,111,294
R-squared	0.622	0.626
Time FE	✓	✓
Appeal X Sector code FE	✓	✓
Trim	1%	5%
KP F-stat	25.21	17.93

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

3.7 Conclusion

In this paper, we study the impact of overturning local authorities' rejection of major residential developments in the UK. Only 20% of developments are rejected, so we understand this sample of projects to be the ones with the most deleterious impact on local neighborhoods. Using property transaction prices as a measure of property value, we are able to rule out a large negative impact of these projects. For some projects, the impact may in fact be positive because they also add to local amenities such as retail shops. This suggests a prevalence of NIMBY-ism, as locals pressure authorities to reject even relatively benign projects. We show that there are more resident comments on rejected projects viz-a-viz accepted projects, which is evidence of NIMBY-ism.

References

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review, 91(5):1369–1401.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2005). The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth. American Economic Review, 95(3):546–579.

Aizer, A. and Doyle Jr, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. Quarterly Journal of Economics, 130(2):759–803.

Akai, N. and Sakata, M. (2002). Fiscal decentralization contributes to economic growth: evidence from state-level cross-section data for the united states. Journal of Urban Economics, 52(1):93–108.

Albouy, D. and Ehrlich, G. (2018). Housing productivity and the social cost of land-use restrictions. Journal of Urban Economics, 107:101–120.

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. Quarterly Journal of Economics, 128(2):155–194.

Alfaro, L., Antras, P., Chor, D., and Conconi, P. (2019). Internalizing Global Value Chains: A Firm-Level Analysis. Journal of Political Economy, 127(2):508–559.

Allen, T. and Donaldson, D. (2020). Persistence and Path Dependence in the Spatial Economy. Working Paper.

Angelucci, C., Meraglia, S., and Voigtländer., N. (2020). How Merchant Towns Shaped Parliaments:

From the Norman Conquest of England to the Great Reform Act. Working Paper.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association, 91(434):444–455.

Antras, P. and Chor, D. (2013). Organizing the Global Value Chain. Econometrica, 81(6):2127–2204.

Atkin, D., Chen, K., and Popov, A. (2019). The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley. Working Paper.

Au, C.-C. and Henderson, J. V. (2006). Are Chinese cities too small? The Review of Economic Studies, 73(3):549–576.

Autor, D. H., Palmer, C. J., and Pathak, P. A. (2014). Housing market spillovers: Evidence from the end of rent control in cambridge, massachusetts. Journal of Political Economy, 122(3):661–717.

Banerjee, A. and Iyer, L. (2005). History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. American Economic Review, 95(4):1190–1213.

Basu, K. (2000). Prelude to Political Economy: A Study of the Social and Political Foundations of Economics. Oxford University Press: New York:.

Baum-Snow, N. (2007). Did highways cause suburbanization? Quarterly Journal of Economics, 122(2):775–805.

Baum-Snow, N. and Marion, J. (2009). The effects of low income housing tax credit developments on neighborhoods. Journal of Public Economics, 93(5-6):654–666.

Bazzi, S., Fiszbein, M., and Gebresilasse, M. (2020). Frontier Culture: The Roots and Persistence of “Rugged Individualism. Econometrica, 88(6):2329–2368.

Beck, T. (2003). Financial dependence and international trade. Review of International Economics, 11:296–316.

Bhuller, M., Dahl, G. B., Løken, K. V., and Mogstad, M. (2020). Incarceration, recidivism, and employment. Journal of Political Economy, 128(4):1269–1324.

Bleakley, H. and Lin, J. (2012). Portage and path dependence. Quarterly Journal of Economics, 127(2):587–644.

Boehm, J. (2020). The Impact of Contract Enforcement Costs on Value Chains and Aggregate Productivity. Review of Economics and Statistics.

Boehm, J. and Oberfield, E. (2020). Misallocation in the Market for Inputs: Enforcement and the Organization of Production. Quarterly Journal of Economics, 135(4):2007–2058.

Calabrese, S., Epple, D., and Romano, R. (2007). On the political economy of zoning. Journal of Public Economics, 91(1-2):25–49.

Cheshire, P., Nathan, M., and Overman, H. (2014). Urban Economics and Urban Policy. Cheltenham: Edward Elgar Publishing.

Cheshire, P. and Sheppard, S. (2002). The welfare economics of land use planning. Journal of Urban Economics, 52(2):242–269.

Chor, D. (2010). Unpacking sources of comparative advantage: a quantitative approach. Journal of International Economics, 82(2):152–167.

Collins, W. J. and Shester, K. L. (2013). Slum clearance and urban renewal in the United States.

American Economic Journal: Applied Economics, 5(1):239–73.

Collinson, R. and Ganong, P. (2018). How do changes in housing voucher design affect rent and neighborhood quality? American Economic Journal: Economic Policy, 10(2):62–89.

Cooter, R. D. (1998). Expressive Law and Economics. Journal of Legal Studies, 27(S2):585–608.

Corkindale, J. (1999). Land Development in the United Kingdom: Private Property Rights and Public Policy Objectives. Environment and Planning A, 31:2053–2070.

Costa, D. L. and Kahn, M. E. (2000). Power Couples: Changes in the Locational Choice of the College Educated, 1940–1990. Quarterly Journal of Economics, 115(4):1287–1315.

Costinot, A. (2009). On the origins of comparative advantage. Journal of International Economics, 77(2):255–264.

Couture, V., Gaubert, C., Handbury, J., and Hurst, E. (2020). Income Growth and the Distributional Effects of Urban Spatial Sorting. Working Paper.

Cunat, A. and Melitz, M. J. (2012). Volatility, labor market flexibility, and the pattern of comparative advantage. Journal of the European Economic Association, 10(2):225–254.

Dahl, G. B., Kostøl, A. R., and Mogstad, M. (2014). Family welfare cultures. Quarterly Journal of Economics, 129(4):1711–1752.

Dale, T. C. (1931). The Inhabitants of London in 1638. Society of Genealogists: London.

Dauth, W., Findeisen, S., Moretti, E., and Suedekum, J. (2018). Matching in Cities. NBER Working Paper No. 25227.

Davis, D. R. and Dingel, J. I. (2020). The comparative advantage of cities. Journal of International Economics, 123:1–27.

Davis, D. R., Dingel, J. I., Monras, J., and Morales, E. (2019). How Segregated Is Urban Consumption? Journal of Political Economy, 127(4):1684–1738.

Davis, D. R. and Weinstein, D. E. (2002). Bones, Bombs, and Break Points: The Geography of Economic Activity. American Economic Review, 95(2):1269–1289.

Davoodi, H. and Zou, H.-f. (1998). Fiscal decentralization and economic growth: A cross-country study. Journal of Urban economics, 43(2):244–257.

Dell, M. (2010). The Persistent Effects of Peru's Mining MITA. Econometrica, 78(6):1863–1903.

Dell, M. and Olken, B. A. (2020). The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java. Review of Economic Studies, 87(1):164–203.

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. American Economic Review, 106(3):479–524.

Diamond, R. and McQuade, T. (2018). Who Wants Affordable Housing in their Backyard? An Equilibrium Analysis of Low Income Property Development. Journal of Political Economy.

Diamond, R., Mcquade, T., and Qian, F. (2019). The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco. American Economic Review, 109(9):3365–3394.

Djankov, S., Porta, R. L., Lopezde-Silanes, F., and Shleifer, A. (2003). Courts. Quarterly Journal of Economics, 118(2):453–517.

Dobbie, W., Goldin, J., and Yang, C. (2018). The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. American Economic Review, 108(2):201–240.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. American Economic Review, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and American Economic Growth: A Market Access Approach. Quarterly Journal of Economics, 131(2):799–858.

Doyle, J. J. J. (2008). Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care. Journal of political Economy, 116(4):746–770.

Duranton, G. and Puga, D. (2004). Handbook of Regional and Urban Economics, Vol 4, chapter Micro-foundations of urban agglomeration economies, pages 2063–2117. North-Holland: Amsterdam.

Duranton, G. and Puga, D. (2012). Urban Growth and Transportation. Review of Economic Studies, 79(4):1407–1440.

Eckert, F., Gvirtz, A., Liang, J., and Peters, M. (2020). A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790. NBER Working Paper 26770.

Ellison, G. and Glaeser, E. L. (1999). The geographic concentration of industry: Does natural advantage explain agglomeration? American Economic Review, 89(2):311–316.

Ellison, G., Glaeser, E. L., and Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. American Economic Review, 100(3):1195–1213.

Engerman, S. L. and Sokoloff, K. L. (1997). How Latin America Fell Behind, chapter Factor

Endowments, Institutions, and Differential Paths of Growth Among New World. Stanford University Press: Stanford.

Epple, D., Romer, T., and Filimon, R. (1988). Community development with endogenous land use controls. Journal of Public Economics, 35(2):133–162.

Faber, B. (2014). Integration and the Periphery: The Unintended Effects of New Highways in a Developing Country. Review of Economic Studies, 81(3):1046–1070.

Field, J. (2008). Reactions and responses to the Great Fire: London and England in the later seventeenth century. PhD thesis, School of Historical Studies, Newcastle University.

Field, J. (2018). London, Londoners and the Great Fire of 1666 – Disaster and Recovery. Routledge: New Jersey.

Fischer, D. H. (1989). Albion's Seed: Four British Folkways in America. Oxford University Press: New York.

Fuchs, V. (1962). Changes in the Location of Manufacturing in the U.S. Since 1929. Yale University Press: New Haven.

Fujita, M. and Thisse, J.-F. (2002). Economics of Agglomeration: Cities, Industrial Location, and Regional Growth. Cambridge University Press: Cambridge.

Galasso, A. and Schankerman, M. (2015). Patents and cumulative innovation: Causal evidence from the courts. Quarterly Journal of Economics, 130(1):317–369.

Garrioch, D. (2016). 1666 and London's Fire History: A Re-Evaluation. The Historical Journal, 59(2):319–338.

Gaubert, C. (2018). Firm Sorting and Agglomeration. American Economic Review, 108(11):3117–3153.

Glaeser, E. L., Gyourko, J., and Saks, R. E. (2005). Why have housing prices gone up? American Economic Review, 95(2):329–333.

Gorodnichenko, Y. and Roland, G. (2017). Culture, Institutions, and the Wealth of Nations. Review of Economics and Statistics, 99(3):402–416.

Greenstone, M. and Gallagher, J. (2008). Does hazardous waste matter? Evidence from the housing market and the superfund program. Quarterly Journal of Economics, 123(3):951–1003.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. Journal of Political Economy, 118(3):536–598.

Greif, A. (1994). Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. Journal of Political Economy, 102(5):912–950.

Guiso, L., Sapienza, P., and Zingales, L. (2016). Long-Term Persistence. Journal of the European Economic Association, 14(6):1401–1436.

Gyourko, J. and Molloy, R. (2015). Regulation and Housing Supply, volume 5. Elsevier B.V., 1 edition.

Hadfield, G. K. and Weingast, B. R. (2012). What Is Law? A Coordination Model of the Characteristics of Legal Order. Journal of Legal Analysis, 4(2):471–514.

Hamilton, B. W. (1975). Zoning and property taxation in a system of local governments. Urban studies, 12(2):205–211.

Hanson, G. and Slaughter, M. J. (2002). Labor-market adjustment in open economies: Evidence from U.S. states. Journal of International Economics, 57(1):3–29.

Head, K. and Mayer, T. (2013). Handbook of International Economics. Vol. 4, chapter Gravity equations: Workhorse, toolkit, and cookbook, pages 131–195. North-Holland: Amsterdam.

Heblich, S., Redding, S. J., and Sturm, D. M. (2020). The Making of the Modern Metropolis: Evidence from London. Quarterly Journal of Economics, 135(4):2059–2133.

Heckman, J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. Econometrica, 73(3):669–738.

Henderson, J. V. (2003). Marshall's scale economies. Journal of Urban Economics, 53(1):1–28.

Hilber, C. A. and Robert-Nicoud, F. (2013). On the origins of land use regulations: Theory and evidence from US metro areas. Journal of Urban Economics, 75(1):29–43.

Hilber, C. A. and Vermeulen, W. (2016). The Impact of Supply Constraints on House Prices in England. Economic Journal, 126(591):358–405.

Holmes, T. J. (1999). Localization of industry and vertical disintegration. Review of Economics and Statistics, 81(2):314–325.

Hornbeck, R. (2010). Barbed wire: Property rights and agricultural development. Quarterly Journal of Economics, 125(2):767–810.

Hornbeck, R. and Keniston, D. (2017). Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872. American Economic Review, 107(6):1365–1398.

Hornbeck, R. and Naidu, S. (2014). When the Levee Breaks: Black Migration and Economic Development in the American South. American Economic Review, 104(3):963–990.

Hornbeck, R. and Rotemberg, M. (2019). Railroads, Reallocation, and the Rise of American Manufacturing. Working Paper.

Hsieh, C. T. and Moretti, E. (2019). Housing constraints and spatial misallocation. American Economic Journal: Macroeconomics, 11(2):1–39.

Humphries, J. E., Mader, N., Tannenbaum, D., and Dijk, W. V. (2018). Does eviction cause poverty? Quasi-experimental evidence from Cook County , IL. Working Paper.

Jedwab, R., Johnson, N. D., and Koyama, M. (2019). Pandemics, Places, and Populations: Evidence from the Black Death. Working Paper.

Jedwab, R. and Moradi, A. (2016). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. Review of Economics and Statistics, 98(2):268–284.

Jha, S. (2015). Financial Asset Holdings and Political Attitudes: Evidence from Revolutionary England. Quarterly Journal of Economics, 130(3):1485–1545.

Jones, P. E. (1966). The Fire Court: Calendar to the Judgments and Decrees of the Court of Judicature Appointed to Determine Differences between Landlords and Tenants as to Rebuilding after the Great Fire, Volume I. The Corporation of London: London.

Jones, P. E. (1970). The Fire Court: Calendar to the Judgments and Decrees of the Court of Judicature Appointed to Determine Differences between Landlords and Tenants as to Rebuilding after the Great Fire, Volume II. The Corporation of London: London.

Kim, S. (1995). Expansion of markets and the geographic distribution of economic activities: The trends in U.S. regional manufacturing structure, 1860–1987. Quarterly Journal of Economics, 110:881–908.

Kline, P. and Moretti, E. (2013). Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. Quarterly Journal of Economics, 129(1):275–331.

Kline, P. and Moretti, E. (2014). People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs. Annual Review of Economics, 6:629–662.

Kostøl, A., Mogstad, M., Setzler, B., et al. (2019). Disability benefits, consumption insurance, and household labor supply. American Economic Review, 109(7):2613–54.

Krugman, P. (1991). History Versus Expectations. Quarterly Journal of Economics, 106(2):651–667.

La Porta, R., de Silanes, F. L., Shleifer, A., and Vishny, R. (1997). Legal Determinants of External Finance. Journal of Finance, 52:1131–1150.

La Porta, R., de Silanes, F. L., Shleifer, A., and Vishny, R. (1998). Law and Finance. Journal of Political Economy, 106:1113–1155.

Levchenko, A. A. (2007). Institutional quality and international trade. Review of Economic Studies, 74(3):791–819.

Libecap, G. D. and Lueck, D. (2011). The Demarcation of Land and the Role of Coordinating Institutions. Journal of Political Economy, 119(3):426–467.

Manova, K. (2008). Credit constraints, equity market liberalizations and international trade. Journal of International Economics, 76(1):33–47.

Manova, K. (2013). Credit constraints, heterogeneous firms, and international trade. Review of Economic Studies, 80(2):711–744.

Marshall, A. (1890). Principles of Economics. Macmillan: London.

McAdams, R. H. (2000). A Focal Point Theory of Expressive Law. Virginia Law Review, 86:1649–1729.

McAdams, R. H. (2005). The Expressive Power of Adjudication. University of Illinois Law Review, 5:1043–1121.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica, 71(6):1695–1725.

Mendez-Chacon, E. and Patten, D. V. (2020). Multinationals, Monopsony and Local Development: Evidence from the United Fruit Company. Working Paper.

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. Review of Economics and Statistics, 90(4):683–701.

Michaels, G. and Rauch, F. (2018). Resetting the Urban Network: 117-2012. Economic Journal, 128:378–412.

Michalopoulos, S. and Pappaioannou, E. (2016). The Long-Run Effects of the Scramble for Africa. American Economic Review, 106(7):1802–1848.

Miguel, E. and Roland, G. (2011). The Long–Run Impact of Bombing Vietnam. Journal of Development Economics, 96(1):1–15.

Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level

Production Functions. American Economic Review, 94(3):656–690.

Myerson, R. B. (2004). Justice, Institutions, and Multiple Equilibria. Chicago Journal of International Law, 5(1):91–107.

Nash, J. F. J. (1950). The Bargaining Problem. Econometrica, 18(2):155–162.

North, D. C. and Thomas, R. P. (1973). The Rise of the Western World. Cambridge University Press: Cambridge.

North, D. C. and Weingast, B. R. (1989). Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England. The Journal of Economic History, 49(4):803–832.

Nunn, N. (2007a). Historical Legacies: A Model Linking Africa's Past to its Current Underdevelopment. Journal of Development Economics, 83(1):157–175.

Nunn, N. (2007b). Relationship-Specificity, Incomplete Contracts and the Pattern of Trade. Quarterly Journal of Economics, 122(2):569–600.

Nunn, N. (2008). The Long-Term Effects of Africa's Slave Trades. Quarterly Journal of Economics, 123(1):139–176.

Nunn, N. and Trefler, D. (2014). Handbook of International Economics Vol. 4., chapter Domestic Institutions as a Source of Comparative Advantage, pages 263–315. North-Holland: Amsterdam.

Nunn, N. and Wantchekon, L. (2011). The Slave Trade and the Origins of Mistrust in Africa. American Economic Review, 101(7):3221–3252.

Ortalo-Magné, F. and Prat, A. (2014). On the political economy of urban growth: Homeownership versus affordability. American Economic Journal: Microeconomics, 6(1 D):154–181.

Ottinger, S. (2020). Immigrants, Industries and Path Dependence. Working Paper.

Oxford Dictionary of National Biography (2004).

Parkhomenko, A. (2020). Local Causes and Aggregate Implications of Land Use Regulation. Working Paper.

Pike, A., Rodríguez-Pose, A., Tomaney, J., Torrisi, G., and Tselios, V. (2012). In search of the 'economic dividend' of devolution: spatial disparities, spatial economic policy, and decentralisation in the uk. Environment and Planning C: Government and Policy, 30(1):10–28.

Porter, R., Gannett, H., and Hunt, W. (1890). Progress of the Nation, including the Map of the Population of 1870. Report on Population of the United States at the Eleventh Census, pages 13–30.

Porter, S. (1996). The Great Fire of London. Sutton Publishing: Gloucestershire.

Quigley, J. M. and Raphael, S. (2005). Regulation and the high cost of housing in California. American Economic Review, 95(2):323–328.

Rappaport, J. and Sachs, J. D. (2003). The United States as a Coastal Nation. Journal of Economic Growth, 8(1):5–46.

Rauch, J. E. (1993). Does History Matter Only When It Matters Little? The Case of City-Industry Location. Quarterly Journal of Economics, 108(3):843–867.

Rauch, J. E. (1999). Networks versus Markets in International Trade. Journal of International Economics, 48(1):7–35.

Reddaway, T. F. (1940). The Rebuilding of London after the Great Fire. Jonathan Cape: London.

Redding, S., Sturm, D., and Wolf, N. (2011). History and Industry Location: Evidence from German Airports. Review of Economics and Statistics, 93(3):814–831.

Richardson, J. (2001). The Annals of London: A Year-by-year Record of a Thousand Years of History. Cassell: London.

Rodríguez-Pose, A. and Bwire, A. (2004). The economic (in) efficiency of devolution. Environment and Planning A, 36(11):1907–1928.

Rossi-Hansberg, E., Sarte, P.-D., and Owens III, R. (2010). Housing Externalities Pierre-Daniel Sarte and Raymond Owens III. The Journal of Poitical Economy, 118(3):1–8.

Sainty, J. (1993). The Judges of England 1272-1990: A List of Judges of the Superior Courts. The Selden Society: London.

Saiz, A. (2010). The Geographic Determinants of Housing Supply. Quarterly Journal of Economics, 125(3):1253–1296.

Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. Review of Economics and Statistics, 88(4):641–658.

Satchell, M., Kitson, P., Newton, G., Shaw-Taylor, L., and Wrigley., T. (2018). 1851 England and Wales census parishes, townships and places [Data Collection]. UK Data Archive: Colchester, Essex.

Sequeira, S., Nunn, N., and Qian, N. (2020). Immigrants and the Making of America. Review of Economic Studies, 87(1):382–419.

Shertzer, A., Twinam, T., and Walsh, R. P. (2018). Zoning and the economic geography of cities. Journal of Urban Economics, 105:20–39.

Siodla, J. (2015). Razing San Francisco: The 1906 Disaster as a Natural Experiment in Urban Redevelopment. Journal of Urban Economics, 89:48–61.

Stansel, D. (2005). Local decentralization and local economic growth: A cross-sectional examination of US metropolitan areas. Journal of Urban Economics, 57(1):55–72.

Tabellini, G. (2010). Culture and Institutions: Economic Development in the Regions of Europe. Journal of the European Economic Association, 8(4):677–716.

The London Gazette (1666). https://www.bl.uk/learning/timeline/item103652.html.

Tian, L. (2019). Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil. Working Paper.

Tidmarsh, J. (2016). The English Fire Courts and the American Right to Civil Jury Trial. University of Chicago Law Review, 83(4):1893–1941.

Turner, F. J. (1893). The Significance of the Frontier in American History. Proceedings of the State Historical Society of Wisconsin.

Turner, M., Haughwout, A., and van der Klaauw, W. (2014). Land Use Regulation and Welfare. Econometrica, 82(4):1341–1403.

Weber, M. (1930). The Protestant Ethic and the Spirit of Capitalism. Routledge: London.

Xie, D., Zou, H.-f., and Davoodi, H. (1999). Fiscal decentralization and economic growth in the United States. Journal of Urban economics, 45(2):228–239.

Zerbe, R. O. and Anderson, C. L. (2001). Culture and Fairness in the Development of Institutions in the California Gold Fields. Journal of Economic History, 61(1):114–143.

Zhang, T. and Zou, H.-f. (1998). Fiscal decentralization, public spending, and economic growth in China. Journal of public economics, 67(2):221–240.

Appendix A: Appendix to Great Expectations: Urban Development in 17th Century London

A.1 The effect of the Fire on quantity

First, did the Fire result in fewer properties being rebuilt? To answer this, we run a difference-in-differences regression where we collapse the data to the parish-level:

$$ln\left(Properties_{jt}\right) = \alpha_j + \delta PostFire_t + \beta Burned_j \times PostFire_t + \gamma' X_{jt} + \epsilon_{jt}$$

$ln\left(Properties_{jt}\right)$ is the log number of properties in parish j in period t. The two periods are before the Fire and after the Fire. $Burned_j$ is an indicator variable that denotes whether the parish experienced damage from the Fire. $PostFire_t$ is an indicator variable for the period after the Fire. X_{jt} is a vector of controls. These include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. These variables are interacted with post-Fire. Broader locations-by-post fixed effects are also included to control for geographical characteristics. Finally, α_j are parish fixed effects. The standard errors are clustered at the parish-level.

Table A.1 presents the results of this regression. The coefficient estimates of β are negative. This is expected as the plague wiped out about a quarter of London's population so we should expect fewer properties to be rebuilt in the immediate aftermath since there are now fewer people to house. The results in column 4 suggest that burned parishes saw a highly statistically significant decrease of around 67.6% properties as compared to unburned parishes. In addition, the reduction in the number of properties is consistent with post-Fire regulations that stipulated that properties needed to be of a certain minimum size.

Table A.1: Effect of Fire on the number of properties

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. Properties)		
Parish Burned X Post Fire	-1.059***	-1.256***	-0.790***	-0.676**
	(0.240)	(0.267)	(0.283)	(0.258)
Observations	140	140	140	140
R-squared	0.205	0.354	0.429	0.460
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Second, did the total number of hearths in the parishes decline after the rebuilding? To answer this, we again run a difference-in-differences regression where we collapse the data to the parish-level:

$$ln\left(Hearths_{jt}\right) = \alpha_j + \delta PostFire_t + \beta Burned_j \times PostFire_t + \gamma' X_{jt} + \epsilon_{jt}$$

$ln\left(Hearths_{jt}\right)$ is the log number of hearths in parish j in period t. The other variables are the same as previously defined and the standard errors are clustered at the parish-level. Table A.2 presents the results from this regression. The results are similar to what happens to the total number of properties being rebuilt after the Fire (Table A.1). In particular, the coefficient estimates are negative.

Table A.2: Effect of Fire on the number of hearths

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. Hearths)		
Parish Burned X Post Fire	-0.866***	-1.043***	-0.643**	-0.518*
	(0.232)	(0.251)	(0.293)	(0.271)
Observations	140	140	140	140
R-squared	0.147	0.324	0.387	0.427
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

A.2 Discussion about Jensen's inequality

Running a regression with a log transformed dependent variable could result in an opposite treatment effect as compared to if we were to run the regression without taking logs. To see this, consider the following stylized example in Table A.3. In this example, the average number of hearths per property and the total number of hearths are higher in parish 2 than in parish 1. However, if we ran a regression using the log of each property's hearths on a parish dummy, we will find that parish 2 on average has fewer log hearths per property.[1]

[1] We would like to thank David Weinstein for providing us with this stylized example.

Table A.3: Stylized example about Jensen's inequality

Parish 1

Property	Hearths	ln(Hearths)
1	10	2.30
2	20	3.00
Parish Average	15	2.65

Parish 2

Property	Hearths	ln(Hearths)
1	5	1.61
2	30	3.40
Parish Average	18	2.51

This stylized example shows the possibility that this could happen but it does not mean that this would definitely happen for other values. Therefore, what we do is to replicate this stylized example using the actual data that we have. In particular, we collapse the data into two groups – burned and unburned parishes. We then compare the differences of the averages (in both logs and without logs) across the burned and unburned groups in the pre- and post-Fire periods. Table A.4 reports the averages from this exercise. It shows us that both a regression without logs and a regression with logs will give us a positive effect. In particular, for the regression without logs we will get a difference-in-differences effect of: $(6.07 - 4.70) - (4.74 - 4.41) = 1.04$. In the regression with logs we get: $(1.80 - 1.55) - (1.56 - 1.48) = 0.17$. Fortunately, the reversal of signs issue does not happen when we use the actual data.

Table A.4: Stylized example using actual data

Unburned

Post-Fire	Parish Average: Hearths	Parish Average: ln(Hearths)
0	4.41	1.48
1	4.74	1.56

Burned

Post-Fire	Parish Average: Hearths	Parish Average: ln(Hearths)
0	4.70	1.55
1	6.07	1.80

The second approach would be to directly run the quality regression without taking logs on the left hand-side variable. This guarantees that the regression will not suffer from Jensen's inequality issues but it comes at the expense of failing the parallel trends assumption and the results being potentially driven by the skewed data. Nevertheless, Table A.5 shows that the estimated coefficient from this regression is positive. Since both the regressions in logs and without logs give us positive coefficient estimates, this should allay the worry that the estimated effect could have different signs when we take logs versus when we do not take logs.

Table A.5: Effect of Fire on the number of hearths per property (no logs)

	(1)	(2)	(3)	(4)
VARIABLES	No. Hearths per Property			
Parish Burned X Post Fire	0.614	0.569	0.407	0.524
	(0.371)	(0.399)	(0.497)	(0.468)
Observations	79,730	79,730	79,730	79,730
R-squared	0.002	0.003	0.003	0.004
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

A.3 Additional figures and tables

Figure A.1: Case characteristics over time

Figure A1.1: Years left in tenancy

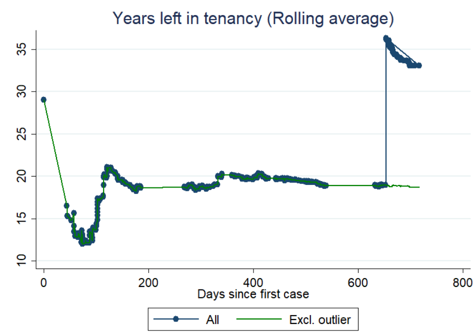


Figure A1.2: Pre-Fire rent

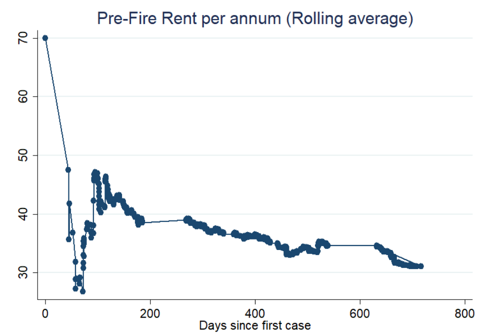


Figure A1.3: Pre-Fire fine

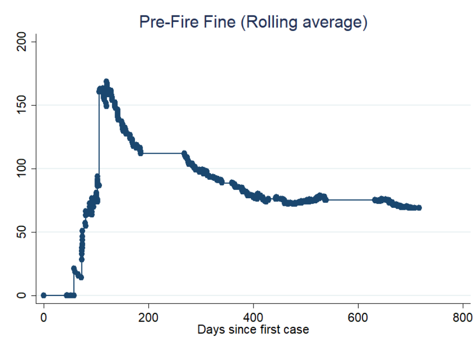


Figure A1.4: Pre-Fire improvements

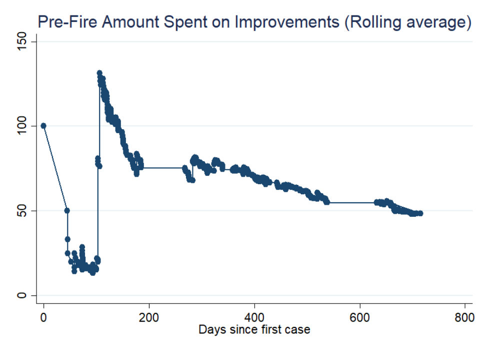


Figure A1.5: Degrees from owner

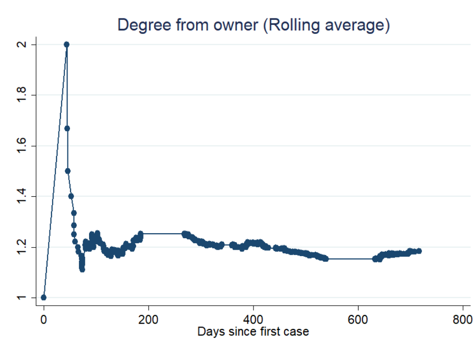


Figure A1.6: Number of parishes

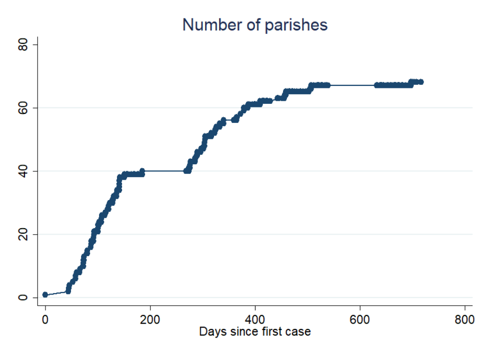


Figure A.2: Binscatter – Effect of Fire on the number of hearths per property (All controls)

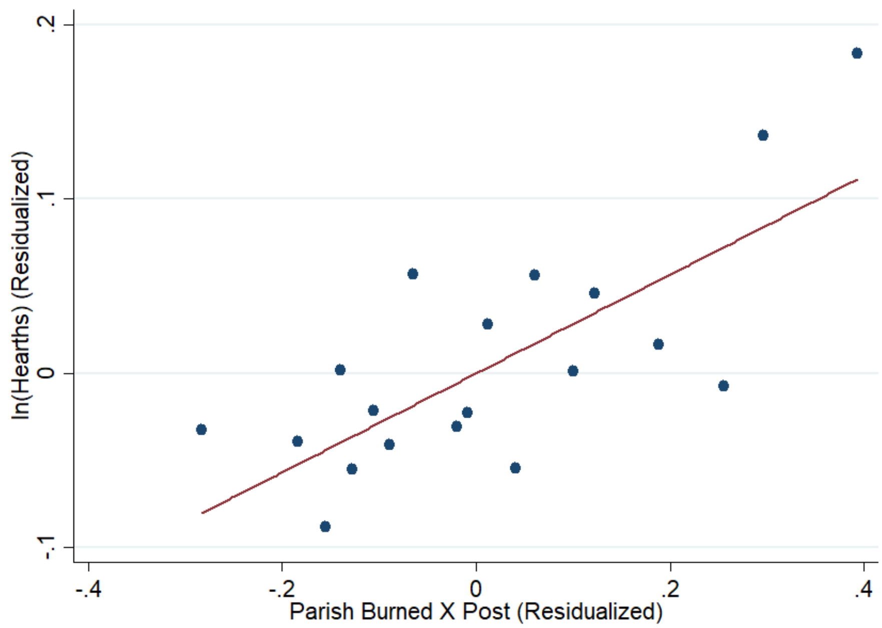


Figure A.3: Binscatter – Effect of legal rulings on the number of hearths per property (All controls)

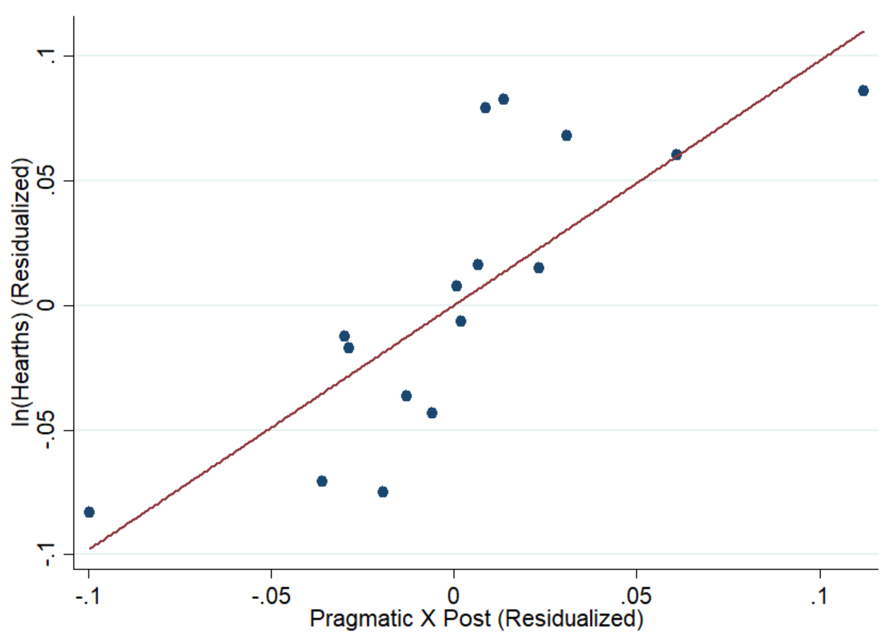


Figure A.4: Binscatter of the first-stage (All controls)

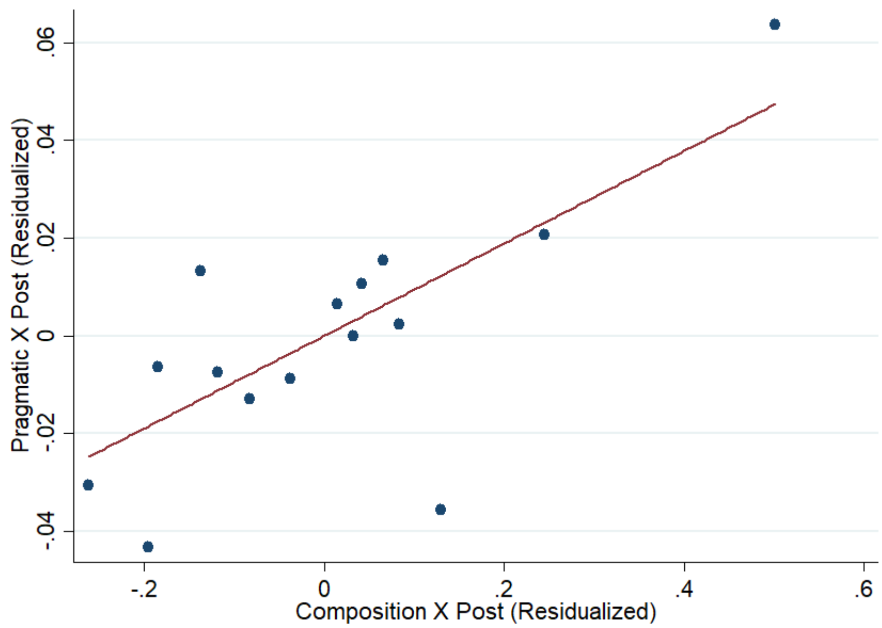


Figure A.5: Binscatter of the reduced-form (All controls)

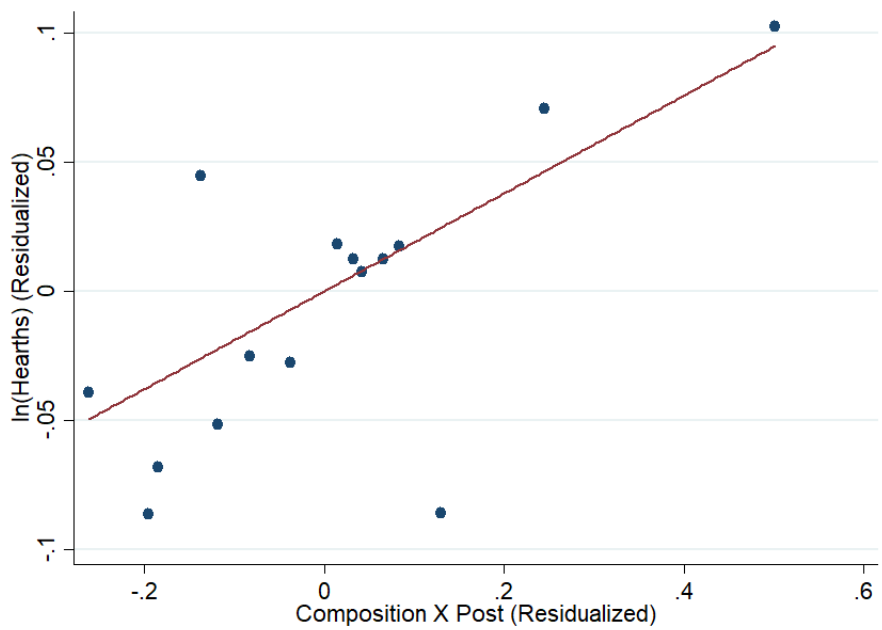


Table A.6: Share of properties in each parish that went to the Fire Court

Parish	Cases	No. properties before Fire	Share
St Botolph Aldersgate	1	3969	0.000
St Giles Cripplegate	1	4967	0.000
St Andrew Holborn	4	1757	.002
All Hallows Staining	1	158	.006
St Antholin Budge Row & St John Walbrook	3	204	.015
St Bartholomew The Less	2	124	.016
St Mary Somerset & St Mary Mounthaw	4	223	.018
St Sepulchre Without Newgate	19	999	.019
All Hallows Barking	9	455	.02
Whitefriars Precinct	4	204	.02
St Bride Fleet Street	34	1614	.021
St Alphage London Wall	4	174	.023
St Martin Ludgate	6	241	.025
Holy Trinity The Less & St Michael Queenhithe	6	226	.027
St Andrew Hubbard & St Mary At Hill	7	255	.027
St Benet Pauls Wharf & St Peter Pauls Wharf	8	298	.027
St Mary Staining & St Michael Wood Street	3	112	.027
St Martin Vintry & St Michael Paternoster Royal	3	105	.029
St Alban Wood Street & St Olave Silver Street	8	257	.031
All Hallows The Great & All Hallows The Less	14	417	.034
St Mary Aldermary & St Thomas Apostle	4	109	.037
St Dunstan In The West	40	1001	.04
St Botolph Billingsgate & St George Botolph Lane	6	148	.041
St Gabriel Fenchurch Street & St Margaret Pattens	6	148	.041
St Swithin London Stone & St Mary Bothaw	7	171	.041
St Dunstan In The East	16	378	.042
Christchurch Newgate Street & St Leonard Foster Lane	24	468	.051
St Magnus The Martyr & St Margaret New Fish Street	12	235	.051
St Peter Le Poer	6	117	.051
St Nicholas Olave & St Nicholas Cole Abbey	6	107	.056
St Matthew Friday Street & St Peter Westcheap	7	117	.06
St Martin Pomeroy & St Olave Old Jewry	7	109	.064
St Michael Le Querne & St Vedast Foster Lane	17	238	.071
St Andrew By The Wardrobe & St Anne Blackfriars	12	167	.072
St Lawrence Jewry & St Mary Magdalen Milk Street	17	231	.074
St Mary Colechurch & St Mildred Poultry	8	108	.074
St Mary Magdalen Old Fish Street	5	68	.074
St Clement Eastcheap & St Martin Orgar	5	65	.077
St Mary Aldermanbury	13	153	.085
St Mary Le Bow & All Hallows Honey Lane & St Pancras Soper Lane	20	194	.103
St Margaret Moses & St Mildred Bread Street	12	107	.112
St Stephen Walbrook & St Benet Sherehog	15	109	.138
St Augustine Watling Street & St Faith Under St Paul	29	203	.143
St Gregory By St Paul	53	364	.146
St Lawrence Pountney & St Mary Abchurch	5	17	.294
All Hallows Bread Street & St John The Evangelist Friday Street	8	27	.296

A.4 Additional results

Table A.7: Comparing parishes before and after the Fire
(burned vs unburned)

	(1)	(2)
VARIABLES	ln(No. hearths)	
Parish Burned	0.101	0.311***
	(0.084)	(0.070)
Observations	42,174	34,919
R-squared	0.141	0.197
Parish controls	✓	✓
Broader location controls	✓	✓
Number of clusters	70	70
Sample	Pre-Fire	Post-Fire

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.8: Effect of Fire on the number of hearths per property (Dropping parishes which merged after the Fire)

VARIABLES	(1)	(2)	(3)	(4)
	ln(No. hearths)			
Parish Burned X Post Fire	0.241**	0.224**	0.222*	0.277**
	(0.105)	(0.109)	(0.129)	(0.120)
Observations	69,466	69,466	69,466	69,466
R-squared	0.007	0.008	0.008	0.010
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	40	40	40	40

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.9: Effect of Fire on the number of hearths per property (Using different control groups - Nearby sample)

VARIABLES	(1)	(2)	(3)	(4)
	ln(No. hearths)			
Parish Burned X Post Fire	0.336***	0.296***	0.327***	0.392***
	(0.094)	(0.080)	(0.116)	(0.121)
Observations	48,103	48,103	48,103	48,103
R-squared	0.013	0.021	0.022	0.025
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	61	61	61	61

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.10: Effect of Fire on the number of hearths per property (Using different control groups - Further away sample)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Parish Burned X Post Fire	0.225**	0.136	0.137	0.236*
	(0.105)	(0.134)	(0.151)	(0.137)
Observations	62,466	62,466	62,466	62,466
R-squared	0.007	0.009	0.010	0.015
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	60	60	60	60

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.11: Effect of Fire on the number of hearths per property (Applying the inverse hyperbolic sine transform to hearths)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Parish Burned X Post Fire	0.124	0.101	0.056	0.086
	(0.112)	(0.134)	(0.159)	(0.159)
Observations	79,730	79,730	79,730	79,730
R-squared	0.002	0.007	0.008	0.009
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.12: Effect of Fire on the number of hearths per property
(Using a Poisson pseudo-likelihood regression)

	(1)	(2)	(3)	(4)
VARIABLES		No. hearths		
Parish Burned X Post Fire	0.129	0.103	0.070	0.101
	(0.088)	(0.108)	(0.129)	(0.121)
Observations	79,730	79,730	79,730	79,730
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.13: Effect of Fire on the number of hearths per property
(Trimming extreme values of the outcome variable)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Parish Burned X Post Fire	0.124**	0.101*	0.086	0.103
	(0.056)	(0.055)	(0.065)	(0.065)
Observations	64,402	64,402	64,402	64,402
R-squared	0.004	0.005	0.005	0.006
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	70	70	70	70

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.14: Comparing parishes before and after the Fire (by legal rulings)

	(1)	(2)
VARIABLES	ln(No. hearths)	
Pragmatic	-0.036	0.769**
	(0.289)	(0.304)
Observations	21,017	10,565
R-squared	0.140	0.196
Parish controls	✓	✓
Broader location controls	✓	✓
Number of clusters	46	46
Sample	Pre-Fire	Post-Fire

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.15: Effect of legal rulings on the number of hearths per property (Adding in other dimensions of the rulings as controls)

	(1)	(2)	(3)	(4)
VARIABLES	ln(No. hearths)			
Pragmatic X Post Fire	1.050***	1.031***	0.926***	0.835***
	(0.294)	(0.306)	(0.249)	(0.224)
Avg. change in tenancy length X Post Fire	0.003	0.000	-0.000	-0.001
	(0.003)	(0.002)	(0.002)	(0.002)
Avg. change in rent X Post Fire	-0.016**	-0.013*	-0.012**	-0.009*
	(0.007)	(0.007)	(0.006)	(0.005)
Observations	31,582	31,582	31,582	31,582
R-squared	0.018	0.026	0.027	0.031
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.16: Effect of legal rulings on the number of hearths per property (Dropping parishes which merged after the Fire)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Pragmatic X Post Fire	2.128***	2.038***	1.917***	2.025***
	(0.657)	(0.249)	(0.304)	(0.409)
Observations	24,384	24,384	24,384	24,384
R-squared	0.024	0.037	0.037	0.037
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	17	17	17	17

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.17: Effect of legal rulings on the number of hearths per property (Applying the inverse hyperbolic sine transform to hearths)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Pragmatic X Post Fire	1.122**	1.091***	1.015***	0.915***
	(0.422)	(0.338)	(0.246)	(0.254)
Observations	32,383	32,383	32,383	32,383
R-squared	0.009	0.017	0.018	0.021
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.18: Effect of legal rulings on the number of hearths per property (Using a Poisson pseudo-likelihood regression)

	(1)	(2)	(3)	(4)
VARIABLES		No. hearths		
Pragmatic X Post Fire	0.921**	0.934***	0.885***	0.784***
	(0.424)	(0.285)	(0.210)	(0.206)
Observations	32,383	32,383	32,383	32,383
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.19: Effect of legal rulings on the number of hearths per property (Trimming extreme values of the outcome variable)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Pragmatic X Post Fire	0.852***	0.709***	0.663***	0.620***
	(0.220)	(0.193)	(0.151)	(0.148)
Observations	25,965	25,965	25,965	25,965
R-squared	0.010	0.015	0.015	0.016
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.20: First-stage by different subsamples

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Pragmatic X Post Fire					
Majority royalist in judging panels X Post	0.135***	0.046	0.038	-0.010	0.179***	0.125**
	(0.048)	(0.049)	(0.062)	(0.060)	(0.032)	(0.051)
Observations	20,737	10,845	1,726	8,419	21,437	19,984
Parish FE	✓	✓	✓	✓	✓	✓
Post FE	✓	✓	✓	✓	✓	✓
Sample	Church not destroyed	Church destroyed	Abutting walls	Within walls	Outside walls	Hearth tercile 1
Number of clusters	11	35	5	33	8	14

	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	Pragmatic X Post Fire					
Majority royalist in judging panels X Post	0.036	0.133**	0.022	-0.096	0.159***	-0.011
	(0.097)	(0.054)	(0.099)	(0.064)	(0.036)	(0.072)
Observations	4,094	7,504	3,094	5,470	23,018	6,690
Parish FE	✓	✓	✓	✓	✓	✓
Post FE	✓	✓	✓	✓	✓	✓
Sample	Hearth tercile 2	Hearth tercile 3	Size tercile 1	Size tercile 2	Size tercile 3	Peers tercile 1
Number of clusters	16	16	19	17	10	21

	(13)	(14)	(15)	(16)	(17)	(18)
VARIABLES	Pragmatic X Post Fire					
Majority royalist in judging panels X Post	0.153***	0.159***	0.073	0.166***	0.062	0.188***
	(0.035)	(0.043)	(0.069)	(0.024)	(0.065)	(0.021)
Observations	17,727	7,165	16,383	15,199	19,135	12,447
Parish FE	✓	✓	✓	✓	✓	✓
Post FE	✓	✓	✓	✓	✓	✓
Sample	Peers tercile 2	Peers tercile 3	Doctors quantile 1	Doctors quantile 2	Military quantile 1	Military quantile 2
Number of clusters	11	14	27	19	39	7

Notes: All regressions include parish FEs and post FE. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.21: Reduced-form – Effect of Royalist majority on the number of hearths per property

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Majority royalist in judging panels X Post	0.270***	0.221**	0.213***	0.189**
	(0.098)	(0.087)	(0.069)	(0.074)
Observations	31,582	31,582	31,582	31,582
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.22: IV – Effect of legal rulings on the number of hearths per property (Adding in other dimensions of the rulings as controls)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Pragmatic X Post Fire	2.198*	2.483**	2.141**	1.951**
	(1.152)	(1.058)	(0.835)	(0.866)
Avg. change in tenancy length X Post Fire	0.004	0.001	0.001	-0.000
	(0.004)	(0.003)	(0.002)	(0.002)
Avg. change in rent X Post Fire	-0.005	-0.004	-0.006	-0.003
	(0.016)	(0.011)	(0.009)	(0.008)
Observations	31,582	31,582	31,582	31,582
R-squared	0.012	0.017	0.022	0.027
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46
KP F-stat	6.427	3.548	8.863	10.63

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.23: IV – Effect of legal rulings on the number of hearths per property (Dropping parishes which merged after the Fire)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Pragmatic X Post Fire	2.065***	1.886***	1.904***	1.945***
	(0.549)	(0.404)	(0.361)	(0.441)
Observations	24,384	24,384	24,384	24,384
R-squared	0.024	0.036	0.037	0.037
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	17	17	17	17
KP F-stat	16.12	7.611	40.30	25.22

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.24: IV – Effect of legal rulings on the number of hearths per property (Applying the inverse hyperbolic sine transform to hearths)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Pragmatic X Post Fire	1.866***	2.271**	1.909**	1.712*
	(0.636)	(0.942)	(0.719)	(0.872)
Observations	32,383	32,383	32,383	32,383
R-squared	0.005	0.010	0.015	0.019
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46
KP F-stat	10.03	3.831	9.195	10.43

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.25: IV – Effect of legal rulings on the number of hearths per property (Trimming extreme values of the outcome variable)

	(1)	(2)	(3)	(4)
VARIABLES		ln(No. hearths)		
Pragmatic X Post Fire	1.663***	1.688***	1.560***	1.490***
	(0.386)	(0.556)	(0.438)	(0.439)
Observations	25,965	25,965	25,965	25,965
R-squared	0.001	0.005	0.008	0.010
Parish FE	✓	✓	✓	✓
Post FE	✓	✓	✓	✓
Parish controls X Post FE		✓	✓	✓
Broader location X Post FE			✓	✓
Pre-fire hearth tercile X Post FE				✓
Number of clusters	46	46	46	46
KP F-stat	12.07	4.716	9.581	10.83

Notes: Parish controls include the number of properties in the parish before the Fire, the share of peers, high-ranking military personnel and doctors living in the parish. Standard errors are clustered at the parish level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Appendix B: Appendix to Made by History: The Spatial Distribution of Manufacturing Industries in the US

B.1 Mathematical Appendix

B.1.1 Microfounding the fixed costs associated with producing in each county

In this Appendix, I show how the higher fixed cost associated with a county having more individualistic culture and institutions can be microfounded using a model of sequential production. The production of final goods entails a continuum of intermediate stages. In each stage, final good producers bargain with suppliers of intermediate inputs. However, due to incomplete contracts, holdup occurs. In counties with more individualistic culture and institutions, this holdup is more severe, leading to a higher fixed cost of production.

Consumers

The consumer's problem is to:

$$\underset{\{q(\omega)\}_{\omega\in\Omega}}{max} \quad U = \left[\int_{\Omega} q(\omega)^{\rho} d\omega\right]^{\frac{1}{\rho}} \quad , \quad \rho \in (0,1)$$

subject to the budget constraint $\int_{\Omega} p(\omega)q(\omega)d\omega \leq Y$, where Y denotes total income. Ω denotes the continuum of possible goods that can be produced and every firm produces a distinct good $\omega \in \Omega$.

Solving this maximization problem yields the optimum quantity demanded of good ω:

$$p(\omega) = q(\omega)^{\rho-1} Y^{1-\rho} P^{\rho}$$

Production function

The production function of each final good firm (i) is:

$$q(i) = z(i)\underbrace{K(i)l(i)^{\gamma-1}}_{Cobb-Douglas}$$

$z(i)$ denotes productivity, $l(i)$ denotes labor input and $K(i)$ denotes intermediate inputs. In particular:

$$K(i) = \left[\int k(i,g)^{\alpha} I(g)dg\right]^{\frac{1}{\alpha}}$$

where $\alpha \in (0,1)$ captures the degree of substitutability among the intermediate inputs. $I(g)$ is an indicator function that equals 1 if input g is produced and intermediate stages $g' < g$ have occured (sequential intermediate inputs).

The timing of decisions for the final good firm is as follows. First, the final good firm bargains sequentially with a continuum $g \in [0,1]$ of suppliers of intermediate inputs $(k(i,g))$. Second, the final good firm chooses the amount of labor $l(i)$. Therefore, in the choosing of labor stage, the cost of the intermediate inputs is a fixed cost as it has been determined in the previous stage. The problem is solved by backward induction.

Choose labor $l(i)$

The final good firm's problem is to:

$$\underset{l(i)}{max}\ p(i)q(i) - wl(i) - f(\phi)$$

subject to the demand for each good $p(i) = q(i)^{\rho-1}Y^{1-\rho}P^{\rho}$ where $q(i) = z(i)K(i)l(i)^{\gamma-1}$. $f(\phi)$ is the fixed cost associated with producing in the county. In particular, $f(\phi) = \beta(\phi)p(i)q(i)$ where $\beta\left(\phi\right)$ is the share of the revenue that accrues to the final good producer's suppliers.

Firm's employment is given by:

$$l(i) = (\gamma - 1)\left(\frac{\rho}{1-\rho}\right)\frac{\pi(i)}{w}$$

The final good firm's optimized profits is thus:

$$\pi(i) = \kappa\left(\frac{z(i)K(i)}{w^{\gamma-1}}\right)^{\frac{\rho}{1-\rho\gamma+\rho}} Y^{\frac{1-\rho}{1-\rho\gamma+\rho}} P^{\frac{\rho}{1-\rho\gamma+\rho}}$$

where $\kappa = [1-\beta(\phi)]\,(\rho(\gamma-1))^{\frac{\rho(\gamma-1)}{1-\rho\gamma+\rho}} - (\rho(\gamma-1))^{\frac{1}{1-\rho\gamma+\rho}}$.

Bargain sequentially with suppliers of intermediate inputs

Intermediate inputs are customized to make them compatible with the needs of the firm producing the final good. There are incomplete contracts since the legal courts are not able to perfectly verify whether inputs are compatible or not. Consequently, given the lack of binding contracts, a holdup problem emerges. I assume that the actual payment to a supplier is negotiated bilaterally only after that stage's input has been produced and the firm has a chance to inspect it. Since the intermdiate input is relationship-specific, the supplier's outside option is zero.

I follow Antras and Chor (2013) and Alfaro et al. (2019) and assume that the final good firm and supplier bargain over how to split the incremental contribution to total revenue generated by the supplier in that stage.

Total revenue for each final good firm is:

$$\begin{aligned} r(i) =& p(i)q(i) \\ =& A\left[\int k(i,g)^{\alpha} I(g)dg\right]^{\frac{\rho}{\alpha}} \end{aligned}$$

where $A = Y^{1-\rho}P^{\rho}z(i)^{\rho}l(i)^{\rho(\gamma-1)}$

Revenue up to stage m is:

$$r(i,m) = A\left[\int_0^m k(i,m)^{\alpha_j}dm\right]^{\frac{\rho_j}{\alpha_j}}$$

Using Leibniz's rule, the incremental contribution to total revenue generated by the supplier in stage m is:

$$\frac{\partial r(i,m)}{\partial m} = \frac{\rho}{\alpha}A^{\frac{\alpha}{\rho}}r(i,m)^{\frac{\rho-\alpha}{\rho}}k(i,m)^{\alpha}$$

where $A = Y^{1-\rho}P^{\rho}z(i)^{\rho}l(i)^{\rho(\gamma-1)}$.

Therefore each supplier's (m) problem is to:

$$\underset{k(i,m)}{max}\ \beta(\phi)\frac{\rho}{\alpha}A^{\frac{\alpha}{\rho}}r(i,m)^{\frac{\rho-\alpha}{\rho}}k(i,m)^{\alpha} - c(m)k(i,m)$$

$\beta(\phi)$ is the share of the incremental contribution that accrues to the supplier and $c(m)$ is the cost to produce one unit of $k(i,m)$. In addition, I make the assumption that $\frac{\partial\beta(\phi)}{\partial\phi} > 0$. This is because when culture and institutions become more individualistic, the holdup problem is more severe and so a greater share of the marginal revenue has to be given to the intermediate good supplier.

Solving this optimization problem gives us the optimal intermediate input (m):

$$k(i,m) = \left(\frac{1-\rho}{1-\alpha}\right)^{\frac{\rho-\alpha}{\alpha(1-\rho)}}\beta(\phi)^{\frac{1}{1-\rho}}A^{\frac{1}{1-\rho}}\rho^{\frac{1}{1-\rho}}\left(\frac{1}{c(m)}\right)\left[\int_0^m\left(\frac{1}{c(m)}\right)^{\frac{\alpha}{1-\alpha}}dm\right]$$

The composite intermediate inputs is thus:

$$K(i) = \left[\int k(i,g)^{\alpha} I(g)dg\right]^{\frac{1}{\alpha}} = \beta(\phi)^{\frac{1}{1-\rho}}\tilde{K}(i)$$

where $\tilde{K}(i) = \left(\frac{1-\rho}{1-\alpha}\right)^{\frac{\rho-\alpha}{\alpha(1-\rho)}} A^{\frac{1}{1-\rho}} \rho^{\frac{1}{1-\rho}} \left\{\int_0^1 \left[\left(\frac{1}{c(m)}\right)^{\frac{1}{1-\alpha}} \left[\int_0^m \left(\frac{1}{c(m)}\right)^{\frac{\alpha}{1-\alpha}} dm\right]^{\frac{\rho-\alpha}{\alpha(1-\rho)}}\right]^{\alpha} I(g)dg\right\}^{\frac{1}{\alpha}}$

Substituting this into the fixed cost:

$$\begin{aligned}
f(\phi) =& \beta(\phi)p(i)q(i) \\
=& \beta(\phi)\left(z(i)K(i)l(i)^{\gamma-1}\right)^{\rho} Y^{1-\rho}P^{\rho} \\
=& \beta(\phi)\left(z(i)\beta(\phi)^{\frac{1}{1-\rho}}\tilde{K}(i)l(i)^{\gamma-1}\right)^{\rho} Y^{1-\rho}P^{\rho} \\
=& \beta(\phi)^{\frac{\rho}{1-\rho}+1}\left(z(i)\tilde{K}(i)l(i)^{\gamma-1}\right)^{\rho} Y_j^{1-\rho}P^{\rho}
\end{aligned}$$

The fixed cost is thus increasing in individualistic culture and institutions (ϕ).

B.1.2 Derivation of Equation 2.1

First, note that profits of a firm in county c can be expressed as:

$$\begin{aligned}
\pi_c(\psi) &= \sum_{r=1}^{\bar{C}} p_{rc}(\psi)y_{rc}(\psi) - \frac{w_c\tau_{rc}}{\psi}y_{rc}(\psi) - f(\phi_c) \\
&= \sum_{r=1}^{\bar{C}} \frac{\sigma}{\sigma-1}\frac{w_c\tau_{rc}}{\psi}y_{rc}(\psi) - \frac{w_c\tau_{rc}}{\psi}y_{rc}(\psi) - f(\phi_c) \\
&= \sum_{r=1}^{\bar{C}} \frac{w_c\tau_{rc}y_{rc}(\psi)}{\psi}\left(\frac{1}{\sigma-1}\right) - f(\phi_c)
\end{aligned}$$

Rearranging:

$$\begin{aligned}
\sum_{r=1}^{\bar{C}} \frac{w_c \tau_{rc} y_{rc}(\psi)}{\psi} \left(\frac{1}{\sigma-1}\right) =& \pi_c(\psi) + f(\phi_c) \\
\sum_{r=1}^{\bar{C}} \frac{w_c \tau_{rc} y_{rc}(\psi)}{\psi} =& (\sigma-1)\pi_c(\psi) + (\sigma-1) f(\phi_c) \\
\sum_{r=1}^{\bar{C}} \frac{w_c \tau_{rc} y_{rc}(\psi)}{\psi} + f(\phi_c) =& (\sigma-1)\pi_c(\psi) + \sigma f(\phi_c)
\end{aligned}$$

Imposing that $\tau_{rc} = \tau_c \geq 1$, this becomes:

$$\sum_{r=1}^{\bar{C}} \frac{\tau_c y_{rc}(\psi)}{\psi} + f(\phi_c) = \frac{\sigma-1}{w_c}\pi_c(\psi) + \frac{\sigma + w_c - 1}{w_c} f(\phi_c)$$

Second, with M_e denoting the number of firms that paid the fixed entry cost, the number of firms producing in the entire country is:

$$M = M_e \left[1 - G(\underline{\psi}_1)\right]$$

Rearranging:

$$M_e = \frac{M}{1 - G(\underline{\psi}_1)}$$

The full employment condition in the country is:

$$
\begin{aligned}
L =&M \int_{\underline{\psi}_1}^{\infty} \sum_{s=1}^{\bar{C}} \sum_{r=1}^{\bar{C}} \left(\frac{\tau_c y_{rs}(\psi)}{\psi} + f(\phi_s) \right) \frac{g(\psi)}{1 - G(\underline{\psi}_1)} d\psi + M_e f_e \\
=&M \int_{\underline{\psi}_1}^{\infty} \sum_{s=1}^{\bar{C}} \left[\frac{\sigma - 1}{w_s} \pi_s(\psi) + \frac{\sigma + w_s - 1}{w_s} f(\phi_s) \right] \frac{g(\psi)}{1 - G(\underline{\psi}_1)} d\psi + \frac{M f_e}{1 - G(\underline{\psi}_1)} \\
=&M \left[(\sigma - 1) \int_{\underline{\psi}_1}^{\infty} \sum_{s=1}^{\bar{C}} \pi_s(\psi) \frac{g(\psi)}{1 - G(\underline{\psi}_1)} d\psi + \sum_{s=1}^{\bar{C}} \frac{\sigma + w_s - 1}{w_s} f(\phi_s) \int_{\underline{\psi}_1}^{\infty} \frac{g(\psi)}{1 - G(\underline{\psi}_1)} d\psi \right] + \frac{M f_e}{1 - G(\underline{\psi}_1)} \\
=&M \left[(\sigma - 1) \tilde{\pi}(\psi) + \sum_{s=1}^{\bar{C}} \frac{\sigma + w_s - 1}{w_s} f(\phi_s) \right] + M \widetilde{\pi}(\psi) \\
=&M \left[\sigma \tilde{\pi}(\psi) + \sum_{s=1}^{\bar{C}} \frac{\sigma + w_s - 1}{w_s} f(\phi_s) \right] \\
=&M \left[\frac{\sigma f_e}{1 - G(\underline{\psi}_1)} + \sum_{s=1}^{\bar{C}} \frac{\sigma + w_s - 1}{w_s} f(\phi_s) \right]
\end{aligned}
$$

Rearranging this, we get the number of firms producing in the entire country:

$$
M = \frac{L}{\left[\frac{\sigma f_e}{\left[1 - G(\underline{\psi}_1)\right]} + \sum_{s=1}^{\bar{C}} \frac{\sigma + w_s - 1}{w_s} f(\phi_s) \right]}
$$

B.1.3 Derivation of $\frac{\partial J_c}{\partial \phi_c}$

First, note that $J_c = \int_{\underline{\psi}_c}^{\infty} \left(\frac{\tau_c y_c(\psi)}{\psi} + f(\phi_c) \right) \frac{g(\psi)}{1 - G(\underline{\psi}_c)} d\psi$

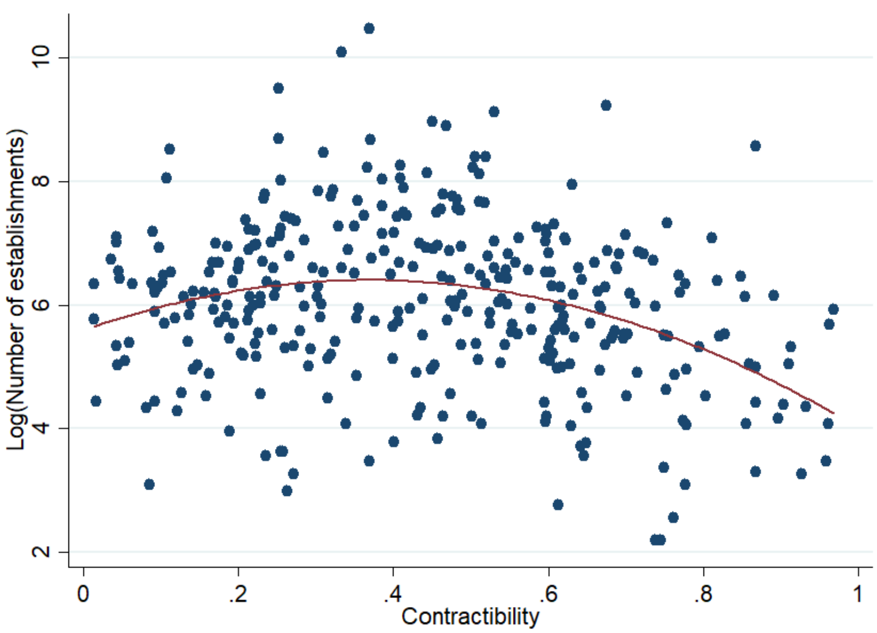
Applying Leibniz's rule:

$$
\begin{aligned}
\frac{\partial J_c}{\partial \phi_c} =& \frac{\partial}{\partial \phi_c} \int_{\underline{\psi}_c}^{\infty} \left(\frac{\tau_c y_c(\psi)}{\psi} + f(\phi_c) \right) \frac{g(\psi)}{1 - G(\underline{\psi}_c)} d\psi \\
=& - \left(\frac{\tau_c y_c(\underline{\psi}_c)}{\underline{\psi}_c} + f(\phi_c) \right) \frac{g(\underline{\psi}_c)}{1 - G(\underline{\psi}_c)} \frac{\partial \underline{\psi}_c}{\partial \phi_c} \\
& + \int_{\underline{\psi}_c}^{\infty} \left(\frac{\partial f(\phi_c)}{\partial \phi_c} \frac{g(\psi)}{1 - G(\underline{\psi}_c)} + \left(\frac{\tau_c y_c(\psi)}{\psi} + f(\phi_c) \right) g(\psi) \left(1 - G(\underline{\psi}_c)\right)^{-2} \frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} \right) d\psi \\
\lesseqgtr & 0
\end{aligned}
$$

since $\frac{\partial f(\phi_c)}{\partial \phi_c} > 0$ and $\frac{\partial G(\underline{\psi}_c)}{\partial \phi_c} = \frac{\partial G(\underline{\psi}_c)}{\partial \underline{\psi}_c} \frac{\partial \underline{\psi}_c}{\partial \phi_c} > 0$.

B.2 Additional figures and tables

Figure B1: Distribution of firms in each industry in 1998



Notes: Each dot represents a distinct industry.

Table B1: Effect of time spent on the frontier on the present-day composition of manufacturing industries (non-coastal counties)

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Outcome: Establishments (Inverse hyperbolic sine transformed)					
Contractibility X Frontier	0.008***	0.006***	0.008***	0.006***	0.006***	0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
R-squared	0.309	0.309	0.309	0.307	0.363	0.432
	B. Outcome: Employment (Inverse hyperbolic sine transformed)					
Contractibility X Frontier	0.022***	0.012**	0.019***	0.014**	0.015**	0.035***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.009)
R-squared	0.236	0.236	0.236	0.234	0.277	0.329
Observations	24,624,380	24,624,380	24,624,380	24,416,140	24,416,140	24,416,140
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table B2: Effect of time spent on the frontier on the present-day composition of manufacturing industries (coastal counties)

	(1)	(2)	(3)	(4)	(5)	(6)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)						
Contractibility X Frontier	0.023***	0.022**	0.023**	0.020**	0.005	0.002
	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)	(0.012)
R-squared	0.479	0.481	0.482	0.478	0.582	0.638
B. Outcome: Employment (Inverse hyperbolic sine transformed)						
Contractibility X Frontier	0.064***	0.045*	0.054**	0.046*	0.023	0.015
	(0.020)	(0.023)	(0.025)	(0.024)	(0.025)	(0.032)
R-squared	0.385	0.387	0.387	0.384	0.470	0.513
Observations	3,271,268	3,271,268	3,271,268	3,243,604	3,243,604	3,243,604
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table B3: Effect of time spent on the frontier on the present-day composition of manufacturing industries (with density controls)

	(1)	(2)	(3)	(4)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)				
Contractibility X Frontier	0.008***	0.007***	0.005**	0.009***
	(0.002)	(0.002)	(0.002)	(0.003)
R-squared	0.358	0.358	0.422	0.491
B. Outcome: Employment (Inverse hyperbolic sine transformed)				
Contractibility X Frontier	0.019***	0.019**	0.010	0.025***
	(0.006)	(0.007)	(0.007)	(0.010)
R-squared	0.272	0.272	0.318	0.370
Observations	27,659,744	27,659,744	27,659,744	27,659,744
Contractibility X Geographical Controls	✓	✓	✓	✓
Contractibility X Historical Controls	✓	✓	✓	✓
Industry Controls X Frontier	✓	✓	✓	✓
Density Controls X Frontier		✓	✓	✓
State X Industy FE			✓	✓
County X 3-digit Industry FE				✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B4: Dropping observations with zeros in the outcome variable

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Outcome: ln(Establishments)					
Contractibility X Frontier	0.039***	0.022**	0.023**	0.029***	0.026***	0.037***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.012)
R-squared	0.478	0.479	0.479	0.476	0.550	0.657
Observations	2,607,556	2,607,556	2,607,556	2,556,327	2,555,960	2,553,883
	B. Outcome: ln(Employment)					
Contractibility X Frontier	0.061***	0.024	0.025	0.036	0.021	0.103***
	(0.019)	(0.020)	(0.021)	(0.022)	(0.021)	(0.030)
R-squared	0.253	0.254	0.254	0.254	0.363	0.501
Observations	2,607,496	2,607,496	2,607,496	2,556,271	2,555,904	2,553,827
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table B5: Using PPML regression to account for the zeros in the outcome variable

	(1)	(2)	(3)	(4)	(5)	(6)
			A. Outcome: Establishments			
Contractibility X Frontier	0.172***	0.113***	0.112***	0.106***	0.090***	0.150***
	(0.034)	(0.031)	(0.032)	(0.033)	(0.030)	(0.036)
			B. Outcome: Employment			
Contractibility X Frontier	0.297***	0.230***	0.221***	0.215***	0.208***	0.255***
	(0.053)	(0.062)	(0.065)	(0.065)	(0.063)	(0.067)
Observations	27,814,765	27,814,765	27,814,765	27,579,545	27,579,545	27,579,545
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table B6: Dropping counties that changed names or FIPS codes and counties that had substantial boundary changes between 1980 and 2016

	(1)	(2)	(3)	(4)	(5)	(6)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)						
Contractibility X Frontier	0.013***	0.009***	0.011***	0.008***	0.008***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
R-squared	0.358	0.359	0.359	0.358	0.424	0.492
B. Outcome: Employment (Inverse hyperbolic sine transformed)						
Contractibility X Frontier	0.035***	0.018***	0.026***	0.018***	0.019***	0.041***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.009)
R-squared	0.274	0.274	0.275	0.273	0.319	0.371
Observations	27,293,519	27,293,519	27,293,519	27,062,707	27,062,707	27,062,707
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table B7: Dropping counties which spent zero years on the frontier

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Outcome: Establishments (Inverse hyperbolic sine transformed)					
Contractibility X Frontier	0.014***	0.014***	0.014***	0.012***	0.011***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
R-squared	0.352	0.353	0.353	0.352	0.404	0.471
	B. Outcome: Employment (Inverse hyperbolic sine transformed)					
Contractibility X Frontier	0.039***	0.035***	0.037***	0.031***	0.027***	0.048***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
R-squared	0.261	0.262	0.262	0.260	0.299	0.351
Observations	22,458,513	22,458,513	22,458,513	22,268,589	22,268,589	22,268,589
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table B8: Dropping industries with non-unique mapping when classifications changed

	(1)	(2)	(3)	(4)	(5)	(6)
A. Outcome: Establishments (Inverse hyperbolic sine transformed)						
Contractibility X Frontier	0.006***	0.004**	0.004**	0.002	0.002	0.007***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
R-squared	0.371	0.372	0.372	0.369	0.432	0.494
B. Outcome: Employment (Inverse hyperbolic sine transformed)						
Contractibility X Frontier	0.021***	0.012**	0.015**	0.008	0.007	0.024***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.009)
R-squared	0.281	0.281	0.281	0.278	0.323	0.376
Observations	17,574,848	17,574,848	17,574,848	17,397,920	17,397,920	17,397,920
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B9: Clustering standard errors based on 60-square mile grid cells

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Outcome: Establishments (Inverse hyperbolic sine transformed)					
Contractibility X Frontier	0.014***	0.009***	0.011***	0.008***	0.009***	0.015***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
R-squared	0.358	0.358	0.359	0.358	0.422	0.490
	B. Outcome: Employment (Inverse hyperbolic sine transformed)					
Contractibility X Frontier	0.036***	0.020**	0.027***	0.019**	0.021***	0.043***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)
R-squared	0.274	0.274	0.274	0.272	0.318	0.370
Observations	27,895,648	27,895,648	27,895,648	27,659,744	27,659,744	27,659,744
Contractibility X Geography		✓	✓	✓	✓	✓
Contractibility X History			✓	✓	✓	✓
Industry Controls X Frontier				✓	✓	✓
State X Industy FE					✓	✓
County X 3-digit Industry FE						✓

Notes: All columns include county, industry and year fixed effects. Geographical controls include latitude, longitude, potential agricultural productivity, area, ruggedness, rainfall risk, distance to the nearest coast, river, lake and mineral deposit. Historical controls include number of years that the county was connected to the railroad by 1890, distance to the nearest conflict with Native Americans, the prevalence of slavery as measured by the population of slaves in each county in 1860 and the share of immigrants in 1910. Industry controls include the log value add of the industry in 1990 and its TFP in 1990. Standard errors clustered at county level. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Appendix C: Appendix to The Inspector Calls: The Effect of Local Land Use Regulations and NIMBY-ism on Housing Prices

C.1 Mathematical Appendix

C.1.1 Proof of Proposition 1

First, we will ignore the PC for both problems and focus just on the IC. We can substitute the FOCs, and apply the envelope theorem to the optimized profits of the honest problem to get:

$$\frac{\partial \pi^h}{\partial \kappa} = \psi'(x-\delta s^*+\kappa)\left[s^* + \omega\theta\underline{s}\mu^*\right] = \psi'(x-\delta s^*+\kappa)\left[s^* + \omega\theta\underline{s}\frac{MP(s^*)}{\omega\theta\underline{s}\delta\psi'(x-\delta s^*+\kappa) - \frac{2s^*}{\alpha} - \frac{\underline{s}(3-2\theta)}{\alpha}}\right]$$

where μ^* is the shadow value of the IC constraint and $MP(s) = -\left[\frac{1}{\alpha} + \delta\psi'(x-\delta s+\kappa)\right]s + P_1(s,\kappa) - as$. We know that $\frac{\partial \pi^h}{\partial \kappa} \geq 0$ because $\mu^* \geq 0$. We can do the same for the optimized profits of the gaming problem:

$$\frac{\partial \pi^g}{\partial \kappa} = \psi'(x-\delta\tilde{s}+\kappa)F\bigg(I(\tilde{s},\kappa)\bigg)\left[\tilde{s} + I(\tilde{s},\kappa)\theta\underline{s}\frac{MP(\tilde{s})}{I(\tilde{s},\kappa)\theta\underline{s}\delta\psi'(x-\delta\tilde{s}+\kappa) - \frac{2\tilde{s}}{\alpha} - \frac{\underline{s}(3-2\theta)}{\alpha}}\right]$$

Define $\underline{\kappa}$ such that $\pi^h(s^*,\underline{\kappa}) = 0$. Since, $\frac{\partial \pi^h}{\partial \kappa} \geq 0$ we know that this is the lowest κ that can satisfy the PC of the honest problem. At $\underline{\kappa}$, $\pi^h(s^*,\underline{\kappa}) \leq \pi^g(\tilde{s},\underline{\kappa})$. If the IC is binding at $\underline{\kappa}$ then $\pi^h(s^*,\underline{\kappa}) < \pi^g(s^u,\underline{\kappa}) \leq \pi^g(\tilde{s},\underline{\kappa})$, where s^u maximizes the unconstrained honest problem, or $MP(s^u) = 0$.

Case 1: $\delta = 0$

$$\frac{\partial \pi^g}{\partial \kappa} = \psi'(x+\kappa)F\bigg(I(\tilde{s},\kappa)\bigg)\left[\tilde{s} + I(\tilde{s},\kappa)\theta\underline{s}\frac{MP(\tilde{s})}{-\frac{2\tilde{s}}{\alpha} - \frac{\underline{s}(3-2\theta)}{\alpha}}\right]$$

$$\leq \psi'(x+\kappa)\left[\tilde{s} + I(\tilde{s},\kappa)\theta\underline{s}\frac{MP(\tilde{s})}{-\frac{2\tilde{s}}{\alpha} - \frac{\underline{s}(3-2\theta)}{\alpha}}\right]$$

To proceed further, note that to satisfy the IC for the gaming problem we need $I(\tilde{s},\kappa) \leq \omega$ and $\tilde{s} \leq s^*$. We can also verify that $\frac{d}{ds}\frac{MP(s)}{-\frac{2s}{\alpha} - \frac{\underline{s}(3-2\theta)}{\alpha}} > 0$. This implies the above term

$$\leq \psi'(x+\kappa)\left[s^* + \omega\theta\underline{s}\frac{MP(s^*)}{-\frac{2s*}{\alpha} - \frac{\underline{s}(3-2\theta)}{\alpha}}\right] = \frac{\partial \pi^h}{\partial \kappa}$$

When $\delta = 0$ we know that the IC for the honest problem will always be satisfied when $\kappa \to \infty$, therefore

$$\lim_{\kappa\to\infty} \pi^h(s^*,\underline{\kappa}) \geq \lim_{\kappa\to\infty} \pi^g(\tilde{s},\underline{\kappa})$$

Since $\pi^h(s^*,\underline{\kappa}) \leq \pi^g(\tilde{s},\underline{\kappa})$, $\frac{\partial \pi^h}{\partial \kappa} \geq 0$ and $\frac{\partial \pi^g}{\partial \kappa} \leq \frac{\partial \pi^h}{\partial \kappa}$, we know that either $\pi^h(s^*,\kappa) = \pi^g(\tilde{s},\kappa) \quad \forall \kappa \in [\underline{\kappa},\infty)$ or the two optimized profits must cross once. Define the value of κ at this crossing as κ^* then we know that $\pi^h(s^*,\kappa) \leq \pi^g(\tilde{s},\kappa) \quad \forall \kappa < \kappa^*$ and $\pi^h(s^*,\kappa) \geq \pi^g(\tilde{s},\kappa) \quad \forall \kappa > \kappa^*$.

Case 2: $\alpha \to \infty$

Under this case, we can verify that

$$\frac{\partial \pi^g}{\partial \kappa} = \frac{F\bigg(I(\tilde{s},\kappa)\bigg)}{\delta}\left[P_1(\tilde{s},\kappa) - \frac{\tilde{s}}{\alpha} - a\tilde{s}\right]$$

$$\frac{\partial \pi^h}{\partial \kappa} = \frac{1}{\delta}\left[P_1(s^*,\kappa) - \frac{s^*}{\alpha} - as^*\right]$$

For the ICs to hold we must have $s^* \leq \tilde{s}$, which implies $\frac{\partial \pi^g}{\partial \kappa} \leq \frac{\partial \pi^h}{\partial \kappa}$. We can then use a similar argument to show a single crossing.

C.1.2 Proof of Proposition 2

Define:

$$H(\kappa, \omega) \equiv \pi^g \left(\tilde{s}; \kappa^*, \omega, \vec{\theta}\right) - \pi^h \left(s^*; \kappa^*, \omega, \vec{\theta}\right) = 0$$

Then, by the implicit function theorem:

$$\frac{\partial \kappa^*}{\partial \omega} = \frac{-H_2(\kappa^*, \omega)}{H_1(\kappa^*, \omega)} = \frac{-\mu^* \beta_\omega(s^*, \omega) I^0}{\frac{\partial \pi^g}{\partial \kappa} - \frac{\partial \pi^h}{\partial \kappa}}$$

From the proof in Appendix C.1.1 we know that at κ^*, $\frac{\partial \pi^h}{\partial \kappa} > \frac{\partial \pi^g}{\partial \kappa} \implies \frac{\partial \kappa^*}{\partial \omega} > 0$

C.2 Additional figures and tables

Table C1: Dynamic difference-in-differences

VARIABLES	(1)	(2)	(3)	(4)
	ln(price)			
Success X Relative Year -2	-0.004	-0.003	-0.004	-0.003
	(0.003)	(0.002)	(0.003)	(0.002)
Success X Relative Year -1	-0.007*	-0.006	-0.007*	-0.006*
	(0.004)	(0.004)	(0.004)	(0.004)
Success X Relative Year 0	-0.011**	-0.011**	-0.010*	-0.010**
	(0.005)	(0.005)	(0.005)	(0.005)
Success X Relative Year 1	-0.011**	-0.011**	-0.012**	-0.011**
	(0.006)	(0.005)	(0.006)	(0.005)
Success X Relative Year 2	-0.016**	-0.017***	-0.016**	-0.016***
	(0.006)	(0.006)	(0.006)	(0.006)
Success X Relative Year 3	-0.023***	-0.019***	-0.022***	-0.019***
	(0.008)	(0.007)	(0.008)	(0.007)
Observations	1,237,943	1,177,861	1,237,706	1,177,683
R-squared	0.576	0.843	0.609	0.856
Time FE	✓	✓	✓	✓
Appeal FE	✓	✓		
Appeal X Postcode sector FE			✓	✓
Controls		✓		✓

Notes: Standard errors clustered at outward code level.
Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table C2: First-stage with controls

	(1)	(2)	(3)
VARIABLES		Success*Postappeal	
Leniency*Postappeal	0.680***	0.664***	0.664***
	(0.136)	(0.135)	(0.135)
Observations	1,236,401	1,176,342	1,176,164
Time FE	✓	✓	✓
Appeal FE	✓	✓	
Appeal X Postcode Sector FE			✓
Controls		✓	✓
KP F-stat	25.15	24.04	24.07

Notes: Standard errors clustered at outward code level.
Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table C3: Effect of overturning the Local Authority's decision (OLS)

	(1)	(2)	(3)	(4)
VARIABLES		ln(price)		
Success*Postappeal	-0.010**	-0.010**	-0.009**	-0.010**
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	1,237,943	1,177,861	1,237,706	1,177,683
R-squared	0.576	0.843	0.609	0.856
Time FE	✓	✓	✓	✓
Appeal FE	✓	✓		
Appeal X Postcode sector FE			✓	✓
Controls		✓		✓

Notes: Standard errors clustered at outward code level.
Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table C4: Effect of overturning the Local Authority's decision (IV)

	(1)	(2)	(3)	(4)
VARIABLES	ln(price)			
Success*Postappeal	0.073**	0.060*	0.072**	0.062*
	(0.035)	(0.033)	(0.034)	(0.033)
Observations	1,236,401	1,176,342	1,236,164	1,176,164
R-squared	-0.002	0.623	-0.003	0.623
Time FE	✓	✓	✓	✓
Appeal FE	✓	✓		
Appeal X Postcode sector FE			✓	✓
Controls		✓		✓
KP F-stat	25.15	24.04	25.08	24.07

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table C5: Effect by distance from appeal site (500m)

	(1)	(2)	(3)	(4)
	OLS		IV	
VARIABLES	ln(price)			
Success*Postappeal	-0.011**	-0.009**	0.039	0.051
	(0.005)	(0.005)	(0.033)	(0.031)
Observations	381,188	360,933	380,811	360,563
R-squared	0.621	0.857	-0.001	0.615
Time FE	✓	✓	✓	✓
Appeal X Postcode sector FE	✓	✓	✓	✓
Controls		✓		✓
Sample	<500m	<500m	<500m	<500m
KP F-stat			31.68	28.96

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table C6: Effect by distance from appeal site (500m to 1km)

	(1)	(2)	(3)	(4)
	OLS		IV	
VARIABLES	ln(price)			
Success*Postappeal	-0.009**	-0.010**	0.095**	0.073*
	(0.004)	(0.004)	(0.042)	(0.038)
Observations	856,381	816,622	855,217	815,474
R-squared	0.626	0.861	-0.004	0.619
Time FE	✓	✓	✓	✓
Appeal X Postcode sector FE	✓	✓	✓	✓
Controls		✓		✓
Sample	500m to 1km	500m to 1km	500m to 1km	500m to 1km
KP F-stat			20.13	19.85

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table C7: Effect by supply shock (small suply shock)

	(1)	(2)	(3)	(4)
	OLS		IV	
VARIABLES	ln(price)			
Success*Postappeal	-0.008	-0.009*	0.081**	0.070**
	(0.005)	(0.005)	(0.036)	(0.035)
Observations	960,928	913,192	960,333	912,619
R-squared	0.626	0.860	-0.003	0.618
Time FE	✓	✓	✓	✓
Appeal X Postcode sector FE	✓	✓	✓	✓
Controls		✓		✓
Sample	Small shock	Small shock	Small shock	Small shock
KP F-stat			25.64	24.70

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table C8: Effect by supply shock (large suply shock)

	(1)	(2)	(3)	(4)
	OLS		IV	
VARIABLES	ln(price)			
Success*Postappeal	-0.004	-0.007	-0.009	-0.011
	(0.005)	(0.005)	(0.100)	(0.083)
Observations	276,778	264,491	275,831	263,545
R-squared	0.518	0.831	0.000	0.646
Time FE	✓	✓	✓	✓
Appeal X Postcode sector FE	✓	✓	✓	✓
Controls		✓		✓
Sample	Large shock	Large shock	Large shock	Large shock
KP F-stat			1.470	1.838

Notes: Standard errors clustered at outward code level.

Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.